Random tensor theory: extending random matrix theory to mixtures of random product states

A. Ambainis

Faculty of Computing, University of Latvia, Riga, Latvia

A. W. Harrow

Department of Mathematics, University of Bristol, Bristol, U.K.

M. B. Hastings

Microsoft Research, Station Q, CNSI Building, University of California, Santa Barbara, CA, 93106

We consider a problem in random matrix theory that is inspired by quantum information theory: determining the largest eigenvalue of a sum of p random product states in $(\mathbb{C}^d)^{\otimes k}$, where k and p/d^k are fixed while $d \to \infty$. When $k = 1$, the Marčenko-Pastur law determines (up to small corrections) not only the largest eigenvalue $((1 + \sqrt{p/d^k})^2)$ but the smallest eigenvalue $(\min(0, 1 - \sqrt{p/d^k})^2)$ and the spectral density in between. We use the method of moments to show that for $k > 1$ the largest eigenvalue is still approximately $(1 + \sqrt{p/d^k})^2$ and the spectral density approaches that of the Marčenko-Pastur law, generalizing the random matrix theory result to the random tensor case. Our bound on the largest eigenvalue has implications both for sampling from a particular heavy-tailed distribution and for a recently proposed quantum data-hiding and correlation-locking scheme due to Leung and Winter.

Since the matrices we consider have neither independent entries nor unitary invariance, we need to develop new techniques for their analysis. The main contribution of this paper is to give three different methods for analyzing mixtures of random product states: a diagrammatic approach based on Gaussian integrals, a combinatorial method that looks at the cycle decompositions of permutations and a recursive method that uses a variant of the Schwinger-Dyson equations.

I. INTRODUCTION AND RELATED WORK

A. Background

A classical problem in probability is to throw p balls into d bins and to observe the maximum occupancy of any bin. If we set the ratio $x = p/d$ to a constant and take d large, this maximum occupancy is $O(\ln d / \ln \ln d)$ with high probability (in fact, this bound is tight, but we will not discuss that here). There are two natural ways to prove this, which we call the large deviation method and the trace method. First, we describe the large deviation method. If the occupancies of the bins are $z_1, \ldots, z_d$, then each z_i is distributed approximately according to a Poisson distribution with parameter x; i.e. $\Pr[z_i = z] \approx x^z/e^x z!$. Choosing $z \gg \ln d / \ln \ln d$ implies that $\Pr[z_i \geq z] \ll 1/d$ for each i. Thus, the union bound implies that with high probability all of the z_i are $\leq O(\ln d / \ln \ln d)$. More generally, the large deviation method proceeds by: (1) representing a bad event (here, maximum occupancy being too large) as the union of many simpler bad events (here, any one z_i being too large), then (2) showing that each individual bad event is very unlikely, and (3) using the union bound to conclude that with high probability none of the bad events occur. This method has been used with great success throughout classical and quantum information theory [17, 18, 22].

This paper will discuss a problem in quantum information theory where the large deviation method fails. We will show how instead a technique called the trace method can be effectively used. For the problem of balls into bins, the trace method starts with the bound $\max z_i^m \leq z_1^m + \ldots + z_d^m$, where m is a large positive integer. Next, we take the expectation of both sides and use convexity to show that

$$\mathbb{E}[\max z_i] \leq (\mathbb{E}[\max z_i^m])^{\frac{1}{m}} \leq d^{\frac{1}{m}} (\mathbb{E}[z_1^m])^{\frac{1}{m}}.$$

Choosing m to minimize the right-hand side can be shown to yield the optimal $\ln d / \ln \ln d + O(1)$ bound for the expected maximum occupancy. In general, this approach is tight up to the factor of $d^{1/m}$.

The quantum analogue of balls-into-bins problem is to choose p random unit vectors $|\varphi_1\rangle, \ldots, |\varphi_p\rangle$ from $\mathbb{C}^d$ and to consider the spectrum of the matrix

$$M_{p,d} = \sum_{s=1}^{p} \varphi_s, \tag{1}$$

where we use φ to denote $|\varphi\rangle\langle\varphi|$. Again we are interested in the regime where $x = p/d$ is fixed and $d \to \infty$. We refer to this case as the "normalized ensemble." We also consider a slightly modified version of the problem in which the states $|\hat{\varphi}_s\rangle$ are drawn from a complex Gaussian distribution with unit variance, so that the expectation of $\langle\hat{\varphi}_s|\hat{\varphi}_s\rangle$ is equal to one. Call the ensemble in the modified problem the "Gaussian ensemble" and define $\hat{M}_{p,d} = \sum_{s=1}^{p}\hat{\varphi}_s$. Note that $\hat{M}_{p,d} = \hat{\Phi}^\dagger\hat{\Phi}$, where $\hat{\Phi} = \sum_{s=1}^{p}|\varphi\rangle\langle s|$ is a $p \times d$ matrix where each entry is an i.i.d. complex Gaussian variable with variance $1/d$. That is, $\hat{\Phi} = \sum_{s=1}^{p}\sum_{j=1}^{d}(a_{s,j} + ib_{s,j})|s\rangle\langle j|$, with $a_{s,j}, b_{s,j}$ i.i.d. real Gaussians each with mean zero and variance $1/2d$.

What we call the Gaussian ensemble is more conventionally known as the Wishart distribution, and has been extensively studied. Additionally, we will see in Section II A that the normalized ensemble is nearly the same as the Gaussian ensemble for large d. In either version of the quantum problem, the larger space from which we draw vectors means fewer collisions than in the discrete classical case. The nonzero part of the spectrum of M has been well studied[12, 32, 33], and it lies almost entirely between $(1 \pm \sqrt{x})^2$ as $d \to \infty$. This can be proven using a variety of techniques. When M is drawn according to the Gaussian ensemble, its spectrum is described by chiral random matrix theory[32, 33]. This follows from the fact that the spectrum of M has the same distribution as the spectrum of the square of the matrix

$$\begin{pmatrix} 0 & \hat{\Phi} \\ \hat{\Phi}^\dagger & 0 \end{pmatrix}, \tag{2}$$

where $\hat{\Phi}$ is defined above. A variety of techniques have been used to compute the spectrum[5, 6, 12, 25]. The ability to use Dyson gas methods, or to perform exact integrals over the unitary group with a Kazakov technique, has allowed even the detailed structure of the eigenvalue spectrum near the edge to be worked out for this chiral random matrix problem.

A large-deviation approach for the $x \ll 1$ case was given in [23, appendix B]. In order to bound the spectrum of $M_{p,d}$, they instead studied the Gram matrix $M'_{p,d} := \Phi\Phi^\dagger$, which has the same spectrum as M. Next they considered $\langle\phi|M'_{p,d}|\phi\rangle$ for a random choice of $|\phi\rangle$. This quantity has expectation 1 and, by Levy's lemma, is within ϵ of its expectation with probability $\geq 1 - \exp(O(d\epsilon^2))$. On the other hand, $|\phi\rangle \in \mathbb{C}^p$, which can be covered by an ϵ-net of size $\exp(O(p\ln 1/\epsilon))$. Thus the entire spectrum of $M'_{p,d}$ (and equivalently $M_{p,d}$) will be contained in $1 \pm O(\epsilon)$ with high probability, where ϵ is a function of x that approaches 0 as $x \to 0$.

In this paper, we consider a variant of the above quantum problem in which none of the techniques described above is directly applicable. We choose our states $|\varphi_s\rangle$ to be product states in $(\mathbb{C}^d)^{\otimes k}$; i.e.

$$|\varphi_s\rangle = |\varphi_s^1\rangle \otimes |\varphi_s^2\rangle \otimes \cdots \otimes |\varphi_s^k\rangle,$$

for $|\varphi_s^1\rangle, \ldots, |\varphi_s^k\rangle \in \mathbb{C}^d$. We choose the individual states $|\varphi_s^a\rangle$ again either uniformly from all unit vectors in $\mathbb{C}^d$ (the normalized product ensemble) or as Gaussian-distributed vectors with $\mathbb{E}[\langle\hat{\varphi}_s^a|\hat{\varphi}_s^a\rangle] = 1$ (the Gaussian product ensemble). The corresponding matrices are $M_{p,d,k}$ and $\hat{M}_{p,d,k}$ respectively. Note that $k = 1$ corresponds to the case considered above; i.e. $M_{p,d,1} = M_{p,d}$ and $\hat{M}_{p,d,1} = \hat{M}_{p,d}$. We are interested in the case when $k > 1$ is fixed. As above, we also fix the parameter $x = p/d^k$, while we take $d \to \infty$. And as above, we would like to show that the spectrum lies almost entirely within the region $(1 \pm \sqrt{x})^2$ with high probability.

However, the Dyson gas and Kazakov techniques[5, 6, 12, 25] that were used for $k = 1$ are not available for $k > 1$, which may be considered a problem of random tensor theory. The difficulty is that we have a matrix with non-i.i.d. entries and with unitary symmetry only within the k subsystems. Furthermore, large-deviation techniques are known to work only in the $x \gg 1$ limits. Here, Ref. [27] can prove that $\|M - xI\| \leq O(\sqrt{xk\log(d)})$ with high probability, which gives the right leading-order behavior only when $x \gg \sqrt{k\log d}$. (The same bound is obtained with different techniques by Ref. [4].) The case when $x \gg 1$ is handled by Ref. [2], which can bound $\|M - xI\| \leq O(\sqrt{x})$ with high probability when $k \leq 2$. (We will discuss this paper further in Section I C 2.)

However, some new techniques will be needed to cover the case when $x \leq O(1)$. Fortunately it turns out that the diagrammatic techniques for $k = 1$ can be modified to work for general k. In Section II, we will use these techniques to obtain an expansion in $1/d$. Second, the large deviation approach of [23] achieves a concentration bound of $\exp(-O(d\epsilon^2))$ which needs to overcome an ϵ-net of size $\exp(O(p\ln(1/\epsilon)))$. This only functions when $p \ll d$, but we would like to take p nearly as large as d^k. One approach when $k = 2$ is to use the fact that $\langle\psi|M_{p,d,k}|\psi\rangle$ exhibits smaller fluctuations when $|\psi\rangle$ is more entangled, and that most states are highly entangled. This technique was used in an unpublished manuscript of Ambainis to prove that $\|M_{p,d,k}\| = O(1)$ with high probability when $p = O(d^2/\operatorname{poly}\ln(d))$. However, the methods in this paper are simpler, more general and achieve stronger bounds.

Our strategy to bound the typical value of the largest eigenvalue of $M_{p,d,k}$ will be to use a trace method: we bound the expectation value of the trace of a high power, denoted m, of $M_{p,d,k}$. This yields an upper bound on $\|M_{p,d,k}\|$

because of the following key inequality

$$\|M_{p,d,k}\|^m \leq \mathrm{tr}(M_{p,d,k}^m). \tag{3}$$

We then proceed to expand $\mathbb{E}[\mathrm{tr}\, M_{p,d,k}^m]$ (which we denote $E_{p,d,k}^m$) as

$$E_{p,d,k}^m = \mathbb{E}[\mathrm{tr}\, M_{p,d,k}^m] = \sum_{s_1=1}^{p} \sum_{s_2=1}^{p} \ldots \sum_{s_m=1}^{p} E_d[s_1, s_2, \ldots, s_m]^k := \sum_{\vec{s} \in [p]^m} E_d[\vec{s}]^k \tag{4}$$

where

$$E_d[\vec{s}] = \mathbb{E}\left[\mathrm{tr}(\varphi_{s_1}\varphi_{s_2}\ldots\varphi_{s_m})\right], \tag{5}$$

and $[p] = \{1, \ldots, p\}$. Similarly we define $\hat{E}_{p,d,k}^m = \mathbb{E}[\mathrm{tr}\, \hat{M}_{p,d,k}^M]$ and $\hat{E}_d[\vec{s}] = \mathbb{E}\left[\mathrm{tr}(\hat{\varphi}_{s_1}\hat{\varphi}_{s_2}\ldots\hat{\varphi}_{s_m})\right]$, and observe that they obey a relation analogous to (4). We also define the normalized traces $e_{p,d,k}^m = d^{-k}E_{p,d,k}^m$ and $\hat{e}_{p,d,k}^m = d^{-k}\hat{E}_{p,d,k}^m$, which will be useful for understanding the eigenvalue density.

The rest of the paper presents three independent proofs that for appropriate choices of m, $E_{p,d,k}^m = (1 + \sqrt{x})^{2m}\exp(\pm o(m))$. This will imply that $\mathbb{E}[\|M_{p,d,k}\|] \leq (1 + \sqrt{x})^2 \pm o(1)$, which we can combine with standard measure concentration results to give tight bounds on the probability that $\|M_{p,d,k}\|$ is far from $(1 + \sqrt{x})^2$. We will also derive nearly matching lower bounds on $E_{p,d,k}^m$ which show us that the limiting spectral density of $M_{p,d,k}$ matches that of the Wishart distribution (a.k.a. the $k = 1$ case). The reason for the multiple proofs is to introduce new techniques to problems in quantum information that are out of reach of the previously used tools. The large-deviation techniques used for the $k = 1$ case have had widely successful applicability to quantum information and we hope that the methods introduced in this paper will be useful in the further exploration of random quantum states and processes. Such random states, unitaries, and measurements play an important role in many area of quantum information such as encoding quantum[1], classical[16, 19], and private[28] information over quantum channels, in other data-hiding schemes[17], in quantum expanders[8, 14], and in general coding protocols[34], among other applications.

The first proof, in Section II, first uses the expectation over the Gaussian ensemble to upper-bound the expectation over the normalized ensemble. Next, it uses Wick's theorem to give a diagrammatic method for calculating the expectations. A particular class of diagrams, called rainbow diagrams, are seen to give the leading order terms. Their contributions to the expectation can be calculated exactly, while for $m \ll d^{1/2k}$, the terms from non-rainbow diagrams are shown to be negligible. In fact, if we define the generating function

$$\hat{G}(x,y) = \sum_{m \geq 0} y^m \hat{e}_{p,d,k}^m, \tag{6}$$

then the methods of Section II can be used to calculate (6) up to $1/d$ corrections. Taking the analytic continuation of $G(x, y)$ gives an estimate of the eigenvalue density across the entire spectrum of $M_{p,d,k}$. More precisely, since we can only calculate the generating function up to $1/d$ corrections, we can use convergence in moments to show that the distribution of eigenvalues weakly converges almost surely (Corollary 6 below) to a limiting distribution. For this limiting distribution, for $x < 1$, the eigenvalue density of $M_{p,d,k}$ vanishes for eigenvalues less than $(1 - \sqrt{x})^2$. However, this calculation, in contrast to the calculation of the largest eigenvalue, only tells us that the fraction of eigenvalues outside $(1 \pm \sqrt{x})^2$ approaches zero with high probability, and cannot rule out the existence of a small number of low eigenvalues.

The second proof, in Section III, is based on representation theory and combinatorics. It first repeatedly applies two simplification rules to $E_d[\vec{s}]$: replacing occurrences of φ_s^2 with φ_s and replacing $\mathbb{E}[\varphi_s]$ with I/d whenever φ_s appears only a single time in a string. Thus $\vec{s}$ is replaced by a (possibly empty) string $\vec{s}'$ with no repeated characters and with no characters occurring only a single time. To analyze $E_d[\vec{s}']$, we express $\mathbb{E}[\varphi^{\otimes n}]$ as a sum over permutations and use elementary arguments to enumerate permutations with a given number of cycles. We find that the dominant contribution (corresponding to rainbow diagrams from Section II) comes from the case when $\vec{s}' = \emptyset$, and also analyze the next leading-order contribution, corresponding to $\vec{s}'$ of the form 1212, 123213, 12343214, 1234543215, etc. Thus we obtain an estimate for $E_{p,d,k}^m$ that is correct up to an $o(1)$ additive approximation.

The third proof, in Section IV, uses the Schwinger-Dyson equations to remove one letter at a time from the string $\vec{s}$. This leads to a simple recursive formula for $e_{p,d,k}^m$ that gives precise estimates.

All three proof techniques can be used to produce explicit calculations of $E_{p,d,k}^m$. Applying them for the first few

values of m yields

$$
\begin{aligned}
E^1_{p,d,k} &= p \\
E^2_{p,d,k} &= p + \frac{(p)_2}{d^k} \\
E^3_{p,d,k} &= p + 3\frac{(p)_2}{d^k} + \frac{(p)_3}{d^{2k}} \\
E^4_{p,d,k} &= p + 6\frac{(p)_2}{d^k} + 6\frac{(p)_3}{d^{2k}} + \frac{(p)_4}{d^{3k}} + 2^k\frac{(p)_2}{d^k(d+1)^k} \\
E^5_{p,d,k} &= p + 10\frac{(p)_2}{d^k} + 20\frac{(p)_3}{d^{2k}} + 10\frac{(p)_4}{d^{3k}} + \frac{(p)_5}{d^{4k}} + 5\cdot 2^k\frac{(p)_2}{d^k(d+1)^k} \\
E^6_{p,d,k} &= p + 15\frac{(p)_2}{d^k} + 50\frac{(p)_3}{d^{2k}} + 50\frac{(p)_4}{d^{3k}} + 15\frac{(p)_5}{d^{4k}} + \frac{(p)_6}{d^{5k}} \\
&\quad + 15\cdot 2^k\frac{(p)_2}{d^k(d+1)^k} + \frac{(p)_2(d+3)^k}{d^{2k}(d+1)^{2k}} + 6^k\frac{(p)_3}{d^k(d+1)^k(d+2)^k},
\end{aligned}
$$

where $(p)_t = p!/(p-t)! = p(p-1)\cdots(p-t+1)$. We see that $O(1)$ (instead of $O(d^k)$) terms start to appear when $m \geq 4$. The combinatorial significance of these will be discussed in Section III B.

B. Statement of results

Our main result is the following theorem.

Theorem 1. Let $\beta_m(x) = \sum_{\ell=1}^m N(m,\ell)x^\ell$, where $N(m,\ell) = \frac{1}{m}\binom{m}{\ell-1}\binom{m}{\ell}$ are known as the Narayana numbers. Then,

$$
\left(1 - \frac{m^2}{p}\right)\beta_m\left(\frac{p}{d^k}\right) \leq \frac{1}{d^k}\mathbb{E}[\mathrm{tr}(M^m_{p,d,k})] \leq \exp\left(\frac{3m^{k+4}}{xd^{1/k}}\right)\beta_m\left(\frac{p}{d^k}\right), \tag{7}
$$

where $\exp(A) := e^A$ and the lower bound holds only when $m < \sqrt{p}$.

Thus, for all $m \geq 1$, $k \geq 1$, $x > 0$ and $p = xd^k$,

$$
\lim_{d\to\infty} e^m_{p,d,k} = \beta_m(p/d^k),
$$

where we have used the notation $e^m_{p,d,k} = \frac{1}{d^k}\mathbb{E}[\mathrm{tr}(M^m_{p,d,k})]$.

Variants of the upper bound are proven separately in each of the next three sections, but the formulation used in the Theorem is proven in Section IV. Since the lower bound is simpler to establish, we prove it only in Section III, although the techniques of Sections II and IV would also give nearly the same bound.

For the data-hiding and correlation-locking scheme proposed in [23], it is important that $\|M\| = 1 + o(1)$ whenever $x = o(1)$. In fact, we will show that $\|M\|$ is very likely to be close to $(1 + \sqrt{x})^2$, just as was previously known for Wishart matrices. First we observe that for large m, $\beta_m(x)$ is roughly $(1 + \sqrt{x})^{2m}$.

Lemma 2.

$$
\frac{x}{2m^2(1+\sqrt{x})^3}(1+\sqrt{x})^{2m} \leq \beta_m(x) \leq (1+\sqrt{x})^{2m} \tag{8}
$$

The proof is deferred to Section I E.

Taking m as large as possible in Theorem 1 gives us tight bounds on the typical behavior of $\|M_{p,d,k}\|$.

Corollary 3. With $M_{p,d,k}$ and x defined as above,

$$
(1+\sqrt{x})^2 - O\left(\frac{\ln(d)}{\sqrt{d}}\right) \leq \mathbb{E}[\|M_{p,d,k}\|] \leq (1+\sqrt{x})^2 + O\left(\frac{\ln(d)}{d^{\frac{1}{2k}}}\right)
$$

and the same bounds hold with $M_{p,d,k}$ replaced by $\hat{M}_{p,d,k}$.

Proof. A weaker version of the upper bound can be established by setting $m \sim d^{1/k(k+4)}$ in

$$\mathbb{E}[\|M_{p,d,k}\|] \leq \mathbb{E}[\|M_{p,d,k}\|^m]^{1/m} \leq d^{\frac{k}{m}}(e_{p,d,k}^m)^{\frac{1}{m}}, \tag{9}$$

where the first inequality is from the convexity of $x \mapsto x^m$. In fact, the version stated here is proven in (42) at the end of Section II.

The lower bound will be proven in Section I E. □

Next, the reason we can focus our analysis on the expected value of $\|M_{p,d,k}\|$ is because $\|M_{p,d,k}\|$ is extremely unlikely to be far from its mean. Using standard measure-concentration arguments (detailed in Section I E), we can prove:

Lemma 4. *For any $\epsilon > 0$,*

$$\Pr(|\|M_{p,d,k}\| - \mathbb{E}[\|M_{p,d,k}\|]| \geq \epsilon) \leq 2\exp(-(d-1)\epsilon^2/k). \tag{10}$$

For any $0 < \epsilon \leq 1$,

$$\Pr\left(\left|\|\hat{M}_{p,d,k}\| - \mathbb{E}[\|M_{p,d,k}\|]\right| \geq \epsilon\mathbb{E}[\|M_{p,d,k}\|] + \delta\right) \leq 2pke^{-\frac{d\epsilon^2}{4k^2}} + 2e^{-\frac{(d-1)\delta^2}{4k}} \tag{11}$$

Combined with Corollary 3 we obtain:

Corollary 5.

$$\Pr\left(|\|M_{p,d,k}\| - \lambda_+|\geq O\left(\frac{\ln d}{d^{1/2k}}\right) + \epsilon\right) \leq 2\exp(-d\epsilon^2/2).$$

A similar, but more cumbersome, bound also exists for $\|\hat{M}_{p,d,k}\|$.

Note that for the $k = 1$ case, the exponent can be replaced by $O(-d\epsilon^{3/2})$, corresponding to typical fluctuations on the order of $O(d^{-2/3})$ [21]. It is plausible that fluctuations of this size would also hold in the $k > 1$ case as well, but we do not attempt to prove that in this paper.

Our asymptotic estimates for $e_{p,d,k}^m$ also imply that the limiting spectral density of $M_{p,d,k}$ is given by the Marčenko-Pastur law, just as was previously known for the $k = 1$ case. Specifically, let $\lambda_1, \ldots, \lambda_R$ be the non-zero eigenvalues of $M_{p,d,k}$, with $R = \text{rank}M_{p,d,k}$. Generically $R = \min(p, d^k)$ and the eigenvalues are all distinct. Define the eigenvalue density to be

$$\rho(\lambda) = \frac{1}{R}\sum_{i=1}^{R}\delta(\lambda_i - \lambda),$$

then

Corollary 6. *In the limit of large d at fixed x, $\rho(\lambda)$ weakly converges almost surely to*

$$\frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{2\pi x\lambda}I(\lambda_- \leq \lambda \leq \lambda_+)$$

for any fixed k and for both the normalized and Gaussian ensembles.

Here $\lambda_\pm = (1 \pm \sqrt{x})^2$ and $I(\lambda_- \leq \lambda \leq \lambda_+) = 1$ if $\lambda_- \leq \lambda \leq \lambda_+$ and 0 otherwise.

This corollary follows from Theorem 1 using standard arguments[10]. We believe, but are unable to prove, that in the $x \leq 1$ case, the probability of any non-zero eigenvalues existing below $\lambda_- - \epsilon$ vanishes for any $\epsilon > 0$ in the limit of large d at fixed x, just as is known when $k = 1$.

C. Applications

1. Data hiding

One of the main motivations for this paper was to analyze the proposed data-hiding and correlation-locking scheme of [23]. In this section, we will briefly review their scheme and explain the applicability of our results.

Suppose that $p = d \log^c(d)$ for some constant $c > 0$, and we consider the k-party state $\rho = \frac{1}{p} \sum_{s=1}^{p} \varphi_s$. We can think of s as a message of $(1 + o(1)) \log d$ bits that is "locked" in the shared state. In [23] it was proved that any LOCC (local operations and classical communication) protocol that uses a constant number of rounds cannot produce an output with a non-negligible amount of mutual information with s, and [23] also proved that the parties cannot recover a non-negligible amount of mutual information with each other that would not be revealed to an eavesdropper on their classical communication so that the state cannot be used to produce a secret key. (They also conjecture that the same bounds hold for an unlimited number of rounds.) However, if $c \log \log(d) + \log(1/\epsilon)$ bits of s are revealed then each party is left with an unknown state from a set of ϵd states in d dimensions. Since these states are randomly chosen, it is possible for each party to correctly identify the remaining bits of s with probability $1 - O(\epsilon)$ [24].

On the other hand, the bounds on the eigenvalues of ρ established by our Corollary 5 imply that the scheme of Ref. [23] can be broken by a separable-once-removed quantum measurement[1]: specifically the measurement given by completing $\{\frac{p}{\|\rho\|} \varphi_s\}_s$ into a valid POVM. We hope that our bounds will also be of use in proving their conjecture about LOCC distinguishability with an unbounded number of rounds. If this conjecture is established then it will imply a dramatic separation between the strengths of LOCC and separable-once-removed quantum operations, and perhaps could be strengthened to separate the strengths of LOCC and separable operations.

2. Sampling from heavy-tailed distributions

A second application of our result is to convex geometry. The matrix M can be thought of as the empirical covariance matrix of a collection of random product vectors. These random product vectors have unit norm, and the distribution has ψ_r norm on the order of $1/\sqrt{d^k}$ iff r satisfies $r \leq 2/k$. Here the ψ_r norm is defined (following [3]) for $r > 0$ and for a scalar random variable X as

$$\|X\|_{\psi_r} = \inf\{C > 0 : \mathbb{E}[\exp(|X|/C)^r] \leq 2\} \tag{12}$$

and for a random vector $|\varphi\rangle$ is defined in terms of its linear forms:

$$\|\varphi\|_{\psi_r} = \sup_{|\alpha\rangle} \|\langle\alpha|\varphi\rangle\|_{\psi_r},$$

where the sup is taken over all unit vectors α. Technically, the ψ_r norm is not a norm for $r < 1$, as it does not satisfy the triangle inequality. To work with an actual norm, we could replace (12) with $\sup_{t \geq 1} \mathbb{E}[|X|^t]^{1/t}/t^{1/r}$, which similarly captures the tail dependence. We mention also that the ψ_2 norm has been called the subgaussian moment and the ψ_1 norm the subexponential moment.

Thm. 3.6 of Ref. [2] proved that when M is a sum of vectors from a distribution with bounded ψ_1 norm and $x \gg 1$ then M is within $O(\sqrt{x} \log x)$ of xI with high probability. And as we have stated, Refs. [4, 27] can prove that $\|M - xI\| \leq O(\sqrt{xk \log(d)})$ with high probability, even without assumptions on r, although the bound is only meaningful when $x \gg \sqrt{k \log d}$. In the case when $x \ll 1$ and $1 \leq r \leq 2$ (i.e. $k \leq 2$), Thm 3.3 of Ref. [3] proved that M is within $O(\sqrt{x} \log^{1/r}(1/x))$ of a rank-p projector. Aubrun has conjectured that their results should hold for $r > 0$ and any distribution on D-dimensional unit vectors with ψ_r norm $\leq O(1/\sqrt{D})$. If true, this would cover the ensembles that we consider.

Thus, our main result bounds the spectrum of M in a setting that is both more general than that of [2, 3] (since we allow general $x > 0$ and $k \geq 1$, implying that $r = 2/k$ can be arbitrarily close to 0) and more specific (since we do not consider only products of uniform random vectors, and not general ensembles with bounded ψ_r norm). Our results can be viewed as evidence in support of Aubrun's conjecture.

D. Notation

For the reader's convenience, we collect here the notation used throughout the paper. This section omits variables that are used only in the section where they are defined.

---

[1] This refers to a POVM (positive operator valued measure) in which all but one of the measurement operators are product operators.

Variable	Definition
d	local dimension of each subsystem.
k	number of subsystems.
p	number of random product states chosen
x	p/d^k.
$|\varphi_s^i\rangle$	unit vector chosen at random from $\mathbb{C}^d$ for $s = 1, \ldots, p$ and $i = 1, \ldots, k$.
$|\hat{\varphi}_s^i\rangle$	Gaussian vector from $\mathbb{C}^d$ with $\mathbb{E}[\langle \varphi_s^i | \varphi_s^i \rangle] = 1$.
φ	$|\varphi\rangle\langle\varphi|$ (for any state $|\varphi\rangle$)
$|\varphi_s\rangle$	$|\varphi_s^1\rangle \otimes \cdots \otimes |\varphi_s^k\rangle$
$|\hat{\varphi}_s\rangle$	$|\hat{\varphi}_s^1\rangle \otimes \cdots \otimes |\hat{\varphi}_s^k\rangle$
$M_{p,d,k}$	$\sum_{s=1}^p \varphi_s$
$\lambda_\pm$	$(1 \pm \sqrt{x})^2$
$E_{p,d,k}^m$	$\mathbb{E}[\operatorname{tr} M_{p,d,k}^m]$
$e_{p,d,k}^m$	$\frac{1}{d^k}\mathbb{E}[\operatorname{tr} M_{p,d,k}^m]$
$E_d[\vec{s}]$	$\mathbb{E}[\operatorname{tr}(\varphi_{s_1} \cdots \varphi_{s_m})]$, where $\vec{s} = (s_1, \ldots, s_m)$
$G(x,y)$	$\sum_{m \geq 0} y^m e_{p,d,k}^m$
$\beta(x)$	$\sum_{\ell=1}^m N(m,\ell)x^\ell$
$N(m,\ell)$	Narayana number: $\frac{1}{m}\binom{m}{\ell-1}\binom{m}{\ell} = \frac{1}{\ell}\binom{m}{\ell-1}\binom{m-1}{\ell-1} = \frac{m!m-1!}{\ell!\ell-1!m-\ell!m-\ell+1!}$ (and $N(0,0) = 1$)
$F(x,y)$	$\sum_{0 < \ell < m < \infty} N(m,\ell)x^\ell y^m$

We also define $|\hat{\varphi}_s\rangle$, $\hat{M}_{p,d,k}$, $\hat{E}_{p,d,k}$, $\hat{G}(x,y)$ and so on by replacing $|\varphi_s^i\rangle$ with $|\hat{\varphi}_s^i\rangle$.

E. Proof of large deviation bounds

In this section we prove Lemma 2, Lemma 4 and the lower bound of Corollary 3. First we review some terminology and basic results from large deviation theory, following Ref. [22]. Consider a set X with an associated measure μ and distance metric D. If $Y \subseteq X$ and $x \in X$ then define $D(x,Y) := \inf_{y \in Y} D(x,y)$. For any $\epsilon \geq 0$ define $Y_\epsilon := \{x \in X : D(x,Y) \leq \epsilon\}$. Now define the concentration function $\alpha_X(\epsilon)$ for $\epsilon \geq 0$ to be

$$\alpha_X(\epsilon) := \max\{1 - \mu(Y_\epsilon) : \mu(Y) \geq 1/2\}.$$

Say that $f : X \to \mathbb{R}$ is η-Lipschitz if $|f(x) - f(y)| \leq \eta D(x,y)$ for any $x, y \in X$. If m is a median value of f (i.e. $\mu(\{x : f(x) \leq m\}) = 1/2$) then we can combine these definitions to obtain the concentration result

$$\mu(\{x : f(x) \geq m + \eta\epsilon\}) \leq \alpha_X(\epsilon). \tag{13}$$

Proposition 1.7 of Ref. [22] proves that (13) also holds when we take $m = \mathbb{E}_\mu[f]$.

Typically we should think of $\alpha_X(\epsilon)$ as decreasingly exponentially with ϵ. For example, Thm 2.3 of [22] proves that $\alpha_{S^{2d-1}}(\epsilon) \leq e^{-(d-1)\epsilon^2}$, where S^{2d-1} denotes the unit sphere in $\mathbb{R}^{2d}$, μ is the uniform measure and we are using the Euclidean distance.

To analyze independent random choices, we define the ℓ_1 direct product $X_{\ell^1}^n$ to be the set of n-tuples $(x_1, \ldots, x_n)$ with distance measure $D_{\ell^1}((x_1, \ldots, x_n), (y_1, \ldots, y_n)) := D(x_1, y_1) + \ldots + D(x_n, y_n)$. Similarly define $X_{\ell^2}^n$ to have distance measure $D_{\ell^2}((x_1, \ldots, x_n), (y_1, \ldots, y_n)) := \sqrt{D(x_1, y_1)^2 + \ldots + D(x_n, y_n)^2}$.

Now we consider the normalized ensemble. Our random matrices are generated by taking pk independent draws from S^{2d-1}, interpreting them as elements of $\mathbb{C}^d$ and then constructing $M_{p,d,k}$ from them. We will model this as the space $((S^{2d-1})_{\ell^2}^p)_{\ell^1}^k$. First, observe that Thm 2.4 of [22] establishes that

$$\alpha_{(S^{2d-1})_{\ell^2}^p}(\epsilon) \leq e^{-(d-1)\epsilon^2}.$$

Next, Propositions 1.14 and 1.15 of [22] imply that

$$\alpha_{((S^{2d-1})_{\ell^2}^p)_{\ell^1}^k}(\epsilon) \leq e^{-(d-1)\epsilon^2/k}.$$

Now we consider the map $f : ((S^{2d-1})_{\ell^2}^p)_{\ell^1}^k \to \mathbb{R}$ that is defined by $f(\{|\varphi_s^i\rangle\}_{\substack{s=1,\ldots,p \\ i=1,\ldots,k}}) = \|M_{p,d,k}\|$, with $M_{p,d,k}$ defined as usual as $M_{p,d,k} = \sum_{s=1}^p \varphi_s^1 \otimes \cdots \otimes \varphi_s^k$. To analyze the Lipschitz constant of f, note that the function $M \to \|M\|$

is 1-Lipschitz if we use the ℓ_2 norm for matrices (i.e. $D(A, B) = \sqrt{\operatorname{tr}(A-B)^\dagger(A-B)}$) [20]. Next, we can use the triangle inequality to show that the defining map from $((S^{2d-1})_{\ell_2}^p)_{\ell_1}^k$ to $M_{p,d,k}$ is also 1-Lipschitz. Thus, f is 1-Lipschitz. Putting this together we obtain the proof of (10).

Next, consider the Gaussian ensemble. any Gaussian vector $|\hat{\varphi}_s^i\rangle$ can be expressed as $|\hat{\varphi}_s^i\rangle = \sqrt{r_{s,i}}|\varphi_s^i\rangle$, where $|\varphi_s^i\rangle$ is a random unit vector in $\mathbb{C}^d$ and $r_{s,i}$ is distributed according to $\chi_{2d}^2/2d$. Here χ_{2d}^2 denotes the chi-squared distribution with $2d$ degrees of freedom; i.e. the sum of the squares of $2d$ independent Gaussians each with unit variance.

The normalization factors are extremely likely to be close to 1. First, for any $t < d$ one can compute

$$\mathbb{E}[e^{tr_{s,i}}] = (1 - t/d)^{-d}.$$

Combining this with Markov's inequality implies that $\Pr[r_{s,i} \geq 1 + \epsilon] = \Pr[e^{tr_{s,i}} \geq e^{t(1+\epsilon)}] \leq (1-t/d)^{-d}e^{-t(1+\epsilon)}$ for any $t > 0$. We will set $t = d\epsilon/(1+\epsilon)$ and then find that

$$\Pr[r_{s,i} \geq 1 + \epsilon] \leq e^{-d(\epsilon - \ln(1+\epsilon))} \leq e^{-\frac{d\epsilon^2}{4}}, \tag{14}$$

where the second inequality holds when $\epsilon \leq 1$. Similarly we can take $t = -d\epsilon/(1-\epsilon)$ to show that

$$\Pr[r_{s,i} \leq 1 - \epsilon] \leq e^{d(\epsilon + \ln(1-\epsilon))} \leq e^{-\frac{d\epsilon^2}{2}}.$$

Now we use the union bound to argue that with high probability none of the $r_{s,i}$ are far from 1. In particular the probability that *any* $r_{s,i}$ differs from 1 by more than ϵ/k is $\leq 2pke^{-d\epsilon^2/4k^2}$.

In the case that all the $r_{s,i}$ are close to 1, we can then obtain the operator inequalities

$$(1-\epsilon)M_{p,d,k} \leq \hat{M}_{p,d,k} \leq (1+2\epsilon)M_{p,d,k}. \tag{15}$$

(For the upper bound we use $(1 + \epsilon/k)^k \leq e^\epsilon \leq 1 + 2\epsilon$ for $\epsilon \leq 1$.) This establishes that $\|\hat{M}_{p,d,k}\|$ is concentrated around the expectation of $\mathbb{E}[\|M_{p,d,k}\|]$, as claimed in (11).

One application of these large deviation bounds is to prove the lower bound in Corollary 3, namely that $(1+\sqrt{x})^2 - O\left(\frac{\ln d}{d^{1/2k}}\right) \leq \mathbb{E}[\|M_{p,d,k}\|]$. First observe that Theorem 1 and Lemma 2 imply that

$$\frac{\left(1 - \frac{m^2}{p}\right)x}{2m^2(1+\sqrt{x})^3}\lambda_+^m \leq e_{p,d,k}^m.$$

On the other hand, $d^{-k}\operatorname{tr} M \leq \|M\|$ and so $e_{p,d,k}^m \leq \mathbb{E}[\|M_{p,d,k}\|^m]$. Define $\mu := \mathbb{E}[\|M_{p,d,k}\|]$. Then

$$
\begin{aligned}
e_{p,d,k}^m &\leq \mathbb{E}[\|M_{p,d,k}\|^m] \\
&= \int_0^\infty d\lambda \Pr[\|M_{p,d,k}\| \geq \lambda]m\lambda^{m-1} && \text{using integration by parts} \\
&\leq \mu^m + m\int_0^\infty d\epsilon(\mu+\epsilon)^{m-1}\Pr[\|M_{p,d,k}\| \geq \mu + \epsilon] \\
&\leq \mu^m + m\int_0^\infty d\epsilon(\mu+\epsilon)^{m-1}e^{-\frac{(d-1)\epsilon^2}{k}} && \text{from (10)} \\
&\leq \mu^m\left(1 + m\int_0^\infty d\epsilon \exp\left((m-1)\epsilon - \frac{(d-1)\epsilon^2}{k}\right)\right) && \text{using } 1 + \epsilon/\mu \leq 1 + \epsilon \leq e^\epsilon \\
&\leq \mu^m\left(1 + m\int_{-\infty}^\infty d\epsilon \exp\left(-\frac{d-1}{k}\left(\epsilon - \frac{k(m-1)}{2(d-1)}\right)^2 + \frac{k^2(m-1)^2}{4(d-1)}\right)\right) && \text{completing the square} \\
&\leq \mu^m\left(1 + m\sqrt{\frac{2\pi k}{d-1}}\exp\left(\frac{k^2(m-1)^2}{4(d-1)}\right)\right) && \text{performing the Gaussian integral}
\end{aligned}
$$

Combining these bounds on $e_{p,d,k}^m$ and taking the mth root we find that

$$\mu \geq \lambda_+\left(\frac{\left(1 - \frac{m^2}{p}\right)x}{2m^2(1+\sqrt{x})^3\left(1 + m\sqrt{\frac{2\pi k}{d-1}}\exp\left(\frac{k^2(m-1)^2}{4(d-1)}\right)\right)}\right)^{\frac{1}{m}}$$

Assuming that $m^2 \leq p/2$ and $m^2 k^2 \leq d$, we find that $\mu \geq \lambda_+ \left(1 - O\left(\frac{\ln(m)}{m}\right)\right) = \lambda_+ \left(1 - O\left(\frac{\ln(d)}{\sqrt{d}}\right)\right)$, which yields the lower bound on $\mathbb{E}[\|M\|]$ stated in Corollary 3. We omit the similar, but more tedious, arguments that can be used to lower-bound $\mathbb{E}[\|\hat{M}\|]$.

We conclude the section with the proof of Lemma 2.

Proof. For the upper bound, note that $N(m, \ell) \leq \binom{m}{\ell}^2 \leq \binom{2m}{2\ell}$ and so

$$\sum_{\ell=1}^{m} N(m, \ell) x^\ell \leq \sum_{\ell'=2}^{2m} \binom{2m}{\ell'} \sqrt{x}^{\ell'} = (1 + \sqrt{x})^{2m}.$$

For the lower bound, first observe that

$$\frac{\binom{2m}{2\ell}}{\binom{m}{\ell}^2} = \frac{2m(2m-1)\cdots(2m-2\ell+1)}{m \cdot m \cdot (m-1) \cdot (m-1) \cdots (m-\ell+1) \cdot (m-\ell+1)} \cdot \frac{\ell!^2}{(2\ell)!} \leq 2^{2\ell} \cdot \frac{\ell!^2}{(2\ell)!} \leq 2\sqrt{\ell} \leq 2\ell.$$

This implies that

$$N(m, \ell) = \frac{\ell}{m(m-\ell+1)} \binom{m}{\ell}^2 \geq \binom{2m}{2\ell}/2m^2. \tag{16}$$

Next, we observe that $\binom{2m}{2\ell+1} \leq \binom{2m}{2\ell} + \binom{2m}{2\ell+2}$, and so by comparing coefficients, we see that

$$\frac{(1 + \sqrt{x})^3}{x} \sum_{\ell=1}^{m} \binom{2m}{2\ell} \geq (1 + \sqrt{x})^{2m}.$$

Combining this with (16) completes the proof of the Lemma. $\qquad \square$

II. APPROACH 1: FEYNMAN DIAGRAMS

A. Reduction to the Gaussian ensemble

We begin by showing how all the moments of the normalized ensemble are always upper-bounded by the moments of the Gaussian ensemble. A similar argument was made in [9, Appendix B]. In both cases, the principle is that Gaussian vectors can be thought of as normalized vectors together with some small fluctuations in their overall norm, and that by convexity the variability in norm can only increase the variance and other higher moments.

Lemma 7. *(a) For all p, d, k, m and all strings $\vec{s} \in [p]^m$,*

$$e^{-\frac{m^2}{2d}} \hat{E}_d[\vec{s}] \leq E_d[\vec{s}] \leq \hat{E}_d[\vec{s}]. \tag{17}$$

(b) For all p, d, k, m,

$$e^{-\frac{m^2 k}{2d}} \hat{E}_{p,d,k}^m \leq E_{p,d,k}^m \leq \hat{E}_{p,d,k}^m. \tag{18}$$

Proof. First note that

$$E_d[\vec{s}] = \left(\prod_{s=1}^{p} \int_{|\varphi_s|^2=1} \mathrm{d}\mu(\varphi_s)\right) \langle \varphi_{s_1}, \varphi_{s_2}\rangle \langle \varphi_{s_2}, \varphi_{s_3}\rangle ... \langle \varphi_{s_m}, \varphi_{s_1}\rangle. \tag{19}$$

where the integral is over $|\varphi_s\rangle \in \mathbb{C}^d$ constrained to $\langle \varphi_s|\varphi_s\rangle = 1$.

Next, for a given choice of $s_1, ..., s_m$, let $\mu_s(s_1, ..., s_m)$ denote the number of times the letter s appears. For example, for $s_1, ..., s_m = 1, 2, 2, 1, 3$ we have $\mu_1 = 2, \mu_2 = 2, \mu_3 = 1$. Then, let us introduce variables r_s and use

$\int_0^\infty \mathrm{d}r_s \exp(-dr_s^2/2)r_s^{2d+2\mu_s-1}\frac{d^{d+\mu_s}}{(d+\mu_s)!} = 1$ to write

$$
\begin{aligned}
E_d[s_1, s_2, ..., s_m] &= \Big(\prod_{s=1}^p \int_{|\varphi_s|^2=1} \mathrm{d}\mu(\varphi_s) \int_0^\infty \mathrm{d}r_s e^{-\frac{dr_s^2}{2}} r_s^{2d-1} \frac{d^{d+\mu_s}}{(d+\mu_s)!}\Big) r_{s_1}\langle\varphi_{s_1},\varphi_{s_2}\rangle r_{s_2}\langle\varphi_{s_2},\varphi_{s_3}\rangle...r_{s_m}\langle\varphi_{s_m},\varphi_{s_1}\rangle &(20) \\
&= \Big(\prod_{s=1}^p \frac{d!d^{\mu_s}}{(d+\mu_s)!}(\frac{d}{2\pi})^d \int \mathrm{d}\hat\varphi_s \exp(-d|\hat\varphi_s|^2/2)\Big)\langle\hat\varphi_{s_1},\hat\varphi_{s_2}\rangle\langle\hat\varphi_{s_2},\hat\varphi_{s_3}\rangle...\langle\hat\varphi_{s_m},\hat\varphi_{s_1}\rangle, \\
&= \Big(\prod_{s=1}^p \frac{d!d^{\mu_s}}{(d+\mu_s)!}\Big)\hat E_d[\vec s]
\end{aligned}
$$

where the integral on the second line is over all $|\hat\varphi_s\rangle \in \mathbb{C}^d$, with

$$|\hat\varphi_s\rangle = r_s|\varphi_s\rangle. \tag{21}$$

Then, since the integral

$$\hat E_d[\vec s] = \Big(\prod_{s=1}^p \Big(\frac{d}{2\pi}\Big)^d \int \mathrm{d}\hat\varphi_s \exp(-d|\hat\varphi_s|^2/2)\Big)\langle\hat\varphi_{s_1},\hat\varphi_{s_2}\rangle\langle\hat\varphi_{s_2},\hat\varphi_{s_3}\rangle...\langle\hat\varphi_{s_m},\hat\varphi_{s_1}\rangle \tag{22}$$

is positive, and

$$1 \geq \prod_{s=1}^p \frac{d!d^{\mu_s}}{(d+\mu_s)!} \geq \frac{1}{\big(1+\frac{1}{d}\big)\cdots\big(1+\frac{m}{d}\big)} \geq e^{-\frac{m(m+1)}{2d}}, \tag{23}$$

we establish (17).

Since $E_{p,d,k}^m$ (resp. $\hat E_{p,d,k}^m$) is a sum over $E_d[\vec s]^m$ (resp. $\hat E_d[\vec s]^m$), each of which is nonnegative, we also obtain (18). This completes the proof of the lemma. $\square$

From now on, we focus on this sum:

$$\hat E_{p,d,k}^m = \sum_{s_1=1}^p \sum_{s_2=1}^p ... \sum_{s_m=1}^p \Big[\Big(\prod_{s=1}^p (\frac{d}{2\pi})^d \int \mathrm{d}\hat\varphi_s \exp(-d|\hat\varphi_s|^2/2)\Big)\langle\hat\varphi_{s_1},\hat\varphi_{s_2}\rangle\langle\hat\varphi_{s_2},\hat\varphi_{s_3}\rangle...\langle\hat\varphi_{s_m},\hat\varphi_{s_1}\rangle\Big]^k \tag{24}$$

We introduce a diagrammatic way of evaluating this sum.

B. Diagrammatics

This section now essentially follows standard techniques in field theory and random matrix theory as used in [12]. The main changes are: first, for $k=1$, our diagrammatic notation will be the same as the usual "double-line" notation, while for $k>1$ we have a different notation with multiple lines. Second, the recursion relation (35) is usually only evaluated at the fixed point where it is referred to as the the "Green's function in the large-d approximation" (or more typically the large-N approximation), while we study how the number of diagrams changes as the number of iterations a is increased in order to verify that the sum of Eq. (34) is convergent. Third, we only have a finite number, $2m$, of vertices, so we are able to control the corrections which are higher order in $1/d$ or $1/p$. In contrast, Ref. [12], for example, considers Green's functions which are sums of diagrams with an arbitrary numbers of vertices.

Integrating Eq. (22) over $\hat\varphi_s$ generates $\prod_s \mu_s!$ diagrams, as shown in Fig. 1. Each diagram is built by starting with one incoming directed line on the left and one outgoing line on the right, with m successive pairs of vertices as shown in Fig. 1(a). We then join the lines coming out of the vertices vertically, joining outgoing lines with incoming lines, to make all possible combinations such that, whenever a pair of lines are joined between the i-th pair of vertices and the j-th pair of vertices, we have $s_i = s_j$. Finally, we join the rightmost outgoing line to the leftmost incoming line; then the resulting diagram forms a number of closed loops. The value of Eq. (22) is equal to the sum over such diagrams of d^{l-m}, where l is the number of closed loops in the diagram. Two example diagrams with closed loops are shown in Fig. 1(b).

Similarly, the sum of Eq. (24) can also be written diagrammatically. There are k incoming lines on the left and k outgoing lines on the right. We have m successive pairs of vertices as shown in Fig. 2(a). Each vertex has now k *pairs* of lines connected vertically: either the solid lines in the pairs are outgoing and the dashed lines are incoming

FIG. 1: a) Vertices for $\hat{E}_d[s_1, s_2]$. b) Example diagrams for $\hat{E}_d[s_1, s_2]$ with $s_1 = s_2$. The diagram on the left has $l = m = 2$ while the diagram on the right (which is also present for $s_1 \neq s_2$) has $l = 1$.
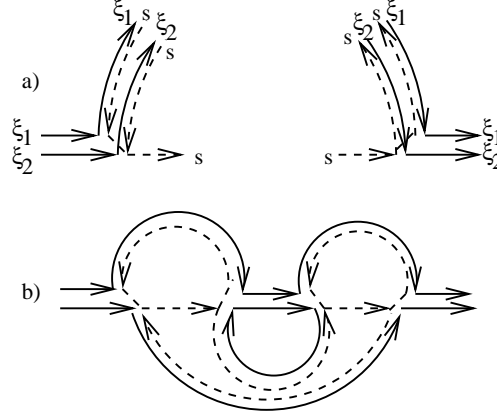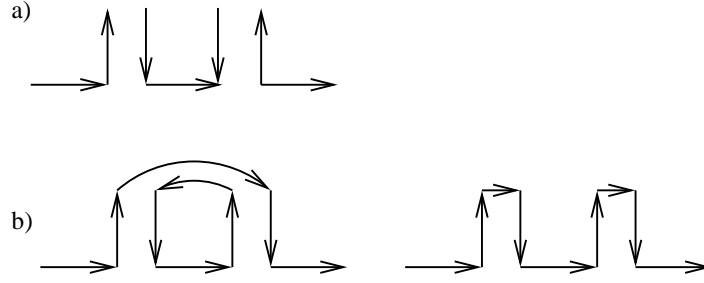

FIG. 2: a) Vertices for diagrams. The left vertex corresponds to $\hat{\Phi}^\dagger$ and the right vertex corresponds to $\hat{\Phi}$. b) An example diagram with $m = 2$ and $k = 2$. There are $l_n = 1 + 2 = 3$ loops of solid lines and $l_p = 1$ disconnected objects of dashed lines.

or vice-versa, depending on whether the vertex has incoming solid lines on the horizontal or outgoing. We label the incoming solid lines by indices $\xi_1, ..., \xi_k \in [d]$, which we refer to as color indices, and then alternately assign to lines along the horizontal axis either a single index of the form $s \in [p]$ for the dashed lines, which we refer to as flavor indices, or k different color indices of the form $\xi_1, ..., \xi_k \in [d]$ for the solid lines. Each of the k lines in a set of k parallel solid lines is also labelled by a "copy index", with the top line labelled as copy 1, the second as copy 2, and so on, up to copy k.

Each of the k pairs of lines coming from a vertex is labelled with a color index ξ and a flavor index s, as well as a copy index. The copy index on a vertical solid line is the same as the copy index of the solid line it connects to on the horizontal, so a given vertex has k distinct copy indices, ranging from $1...k$. Each diagram consists of a way of joining different pairs of vertical lines, subject to the rule that when we join two vertical lines, both have the same copy index; thus, if a given vertical line comes from the k'-th row, $1 \leq k' \leq k$, then it must join to a line which also comes from the k'-th row.

The value of a diagram is equal to d^{-m} times the number of possible assignments of values to the indices, such that whenever two lines are joined they have the same indices. The solid lines break up into some number l_n different closed loops; again, when counting the number of closed loops, we join the solid lines leaving on the right-hand side of the diagram to those entering on the left-hand side of the diagram. Since all solid lines in a loop have the same copy index, we have $l_n = l_{n,1} + l_{n,2} + ... + l_{n,k}$, where $l_{n,k'}$ is the number of loops of solid lines with copy index k'. The dashed lines s come together in vertices where k different lines meet. Let l_p denote the number of different disconnected sets of dashed lines. Then, the value of a diagram is equal to

$$d^{-mk} d^{l_n} p^{l_p}. \tag{25}$$

Note, we refer to disconnected sets of lines in the case of dashed lines; this is because multiple lines meet at a single vertex; for $k = 1$ these sets just become loops. An example diagram is shown in Fig. 2(b) for $k = 2$.

Let $c_m^k(l_n, l_p)$ equal the number of diagrams with given l_n, l_p for given m, k. Then,

$$\hat{E}_{p,d,k}^m = \sum_{l_n \geq 1} \sum_{l_p \geq 1} c_m^k(l_n, l_p) d^{-mk} d^{l_n} p^{l_p}. \tag{26}$$
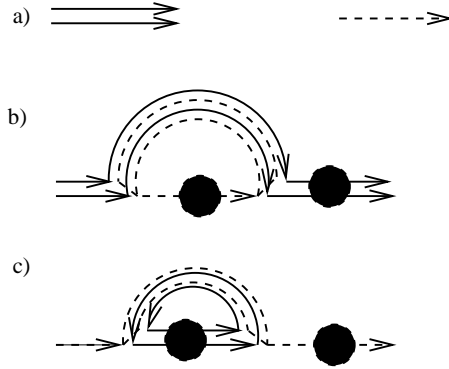
FIG. 3: Iterative construction of rainbow diagrams for $k = 2$. The solid lines with a filled circle denotes any open rainbow diagram as does the dashed line with a filled circle.
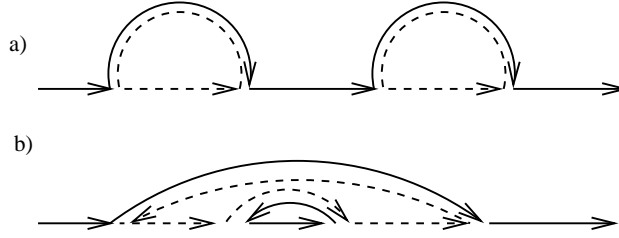

FIG. 4: (a,b) Rainbow diagrams which require two iterations for $k = 1$.

C.   Rainbow Diagrams

An important set of diagrams are the so-called "rainbow diagrams", which will be the dominant contributions to the sum (26). We define these rainbow diagrams with the following iterative construction.

We define a group of k solid lines or a single dashed line to be an open rainbow diagram as shown in Fig. 3(a). We also define any diagram which can be contructed as in Fig. 3(b,c) to be a open rainbow diagram, where the k solid lines or one dashed line with a filled circle may be replaced by any open rainbow diagram. We say that the rainbow diagrams in Fig. 3(b,c) has solid and dashed external lines respectively.

In general all open rainbow diagrams can be constructed from the iterative process described in Fig. 3(b,c), with one "iteration" consisting of replacing one of the filled circles in Fig. 3(b,c) with one of the diagrams in Fig. 3. The diagrams in Fig. 3(a) require zero iterations, and each iteration adds one vertex. For example, in Fig. 4(a,b) we show the two diagrams with solid external lines which require two iterations to construct for $k = 1$. We define a rainbow diagram to be any open rainbow diagram where we assume that the right outgoing edge and left incoming edge are solid lines and are connected.

D.   Combinatorics of Diagrams and Number of Loops

We now go through several claims about the various diagrams. The goal will be to count the number of diagrams for given l_n, l_p. First, we claim that for the rainbow diagrams

$$l_n + kl_p = (m + 1)k, \tag{27}$$

as may be directly verified from the construction. Next, we claim that for any diagram

$$l_n + kl_p \leq (m + 1)k. \tag{28}$$

From Eq. (27) the rainbow diagrams saturate this bound (28). We claim that it suffices to show Eq. (28) for $k = 1$ in order to show Eq. (28) for all k. To see this, consider any diagram for $k > 1$. Without loss of generality, suppose $l_{n,1} \geq l_{n,k'}$ for all $1 \leq k' \leq k$. Then, $l_n + kl_p \leq k(l_{n,1} + p)$. We then remove all the solid lines on the horizontal with copy indices $2...k$, as well as all pairs of lines coming from a vertex with copy indices $2...k$. Having done this, both

the solid and the dashed lines form closed loops, since only two dashed lines meet at each vertex. The new diagram is a diagram with $k = 1$. The number of loops of solid lines is $l_{n,1}$, while the number of loops of dashed lines in the new diagram, l'_p, is greater than or equal to l_p since we have removed dashed lines from the diagram. Thus, if we can show Eq. (28) for $k = 1$, it will follow that $l_{n,1} + l'_p \leq (m + 1)$ and so $l_n + k l_p \leq (m + 1)k$.

To show Eq. (28) for $k = 1$, we take the given diagram, and make the replacement as shown between the left and right half of Fig. 5(a): first we straighten the diagram out as shown in the middle of Fig. 5(a), then we replace the double line by a wavy line connected the solid and dashed lines. Finally, we take the point where the solid line leaves the right-hand side of the diagram and connects to the solid line entering the left-hand side and put a single dot on this point for reference later as shown in Fig. 5(b,c). Having done this, the diagram consists of closed loops of solid or dashed lines, with wavy lines that connect solid to dashed lines, and with one of the closed loops of solid lines having a dot on it at one point.

This procedure gives an injective mapping from diagrams written as in Fig. 2 to diagrams written as in Fig. 5. However, this mapping is not invertible; when we undo the procedure of Fig. 5(a), we find that some diagrams can only be written as in Fig. 2 if there are two or more horizontal lines. The diagrams which are the result of applying this procedure to a diagram as in Fig. 2 with only one horizontal line are those that are referred to in field theory as contributions to the "quenched average," while the sum of all diagrams, including those not in the quenched average, is referred to as the "annealed average". To determine if a diagram is a contribution to the quenched average, start at the dot and then follow the line in the direction of the arrow, crossing along a wavy line every time it is encountered, and continuing to follow solid and dashed lines in the direction of the arrow, and continuing to cross every wavy line encountered. Then, a diagram is a contribution to the quenched average if and only if following the lines in this manner causes one to traverse the entire diagram before returning to the starting point, while traversing wavy lines in both directions. As an example, consider the diagram of Fig. 5(c): this diagram is not a contribution to the quenched average, as can be seen by traversing the diagram, or by re-drawing the diagram as in Fig. 5(d) which requires two horizontal solid lines[2]. If a diagram is a contribution to the quenched average, then traversing the diagram in this order (following solid, dashed, and wavy lines as above) corresponds to traversing the diagram writen as in Fig. 2 from left to right.

Since all diagrams are positive, we can bound the sum of diagrams which are contributions to the quenched average by bounding the sum of all diagrams as in Fig. 5. The number of wavy lines is equal to m. The diagram is connected so therefore the number of solid plus dashed loops, which is equal to $l_{n,1} + l'_p$, is at most equal to the number of wavy lines plus one. Therefore, Eq. (28) follows. From this construction, the way to saturate Eq. (28) is to make a diagram which is a tree whose nodes are closed loops of dashed and solid lines and whose edges are wavy lines; that is, a diagram such that the removal of any wavy line breaks the diagram into two disconnected pieces. These trees are the same as the rainbow diagrams above. In Fig. 5(b) we show the two different trees which correspond to the rainbow diagrams of Fig. 4.

Next, we consider the diagrams which are not rainbow diagrams. First, we consider the case $k = 1$. Let $d = m + 1 - l_n - l_p \geq 0$. If $d > 0$, then the diagram is not a rainbow diagram. However, if $d > 0$, using the construction above with closed loops connected by wavy lines, there are only $l_n + l_p$ loops connected by more than $l_n + l_p - 1$ wavy lines; this implies that the diagram is not a tree (using the notation of Fig. 5) or a rainbow diagram (using the notation of Fig. 4), and hence it is possible to remove d different lines and arrive at a diagram which is a rainbow diagram. Thus, all diagrams with $2m$ vertices and $d > 0$ can be formed by taking rainbow diagrams with $2m - d$ vertices and adding d wavy lines; these wavy lines can be added in at most $[m(m-1)]^d$ different ways. Thus, for $k = 1$ we have

$$m + 1 - l_n - l_p = d > 0 \ \rightarrow \ c_m^1(l_n, l_p) \ \leq \ c_{m-d}^1(l_n, l_p) m^{2d} \tag{29}$$

We now consider the number of diagrams which are not rainbow diagrams for $k > 1$. We consider all diagrams, including those which contribute to the annealed average, but we write the diagrams as in Fig. 2, possibly using multiple horizontal lines. Consider first a restricted class of diagrams: those diagrams for which, for every vertex with k pairs of lines leaving the vertex, all k of those pairs of lines connect with pairs of lines at the *same* vertex. This is not the case for, for example, the diagram of Fig. 2(b), where of the two pairs of lines leaving the leftmost vertex, the top pair reconnects at the second vertex from the left, while the bottom pair reconnects at the rightmost vertex. However, for a diagram in this restricted class, the counting of diagrams is exactly the same as in the case $k = 1$, since the diagrams in this restricted class are in one-to-one correspondence with those for $k = 1$. So, the number of
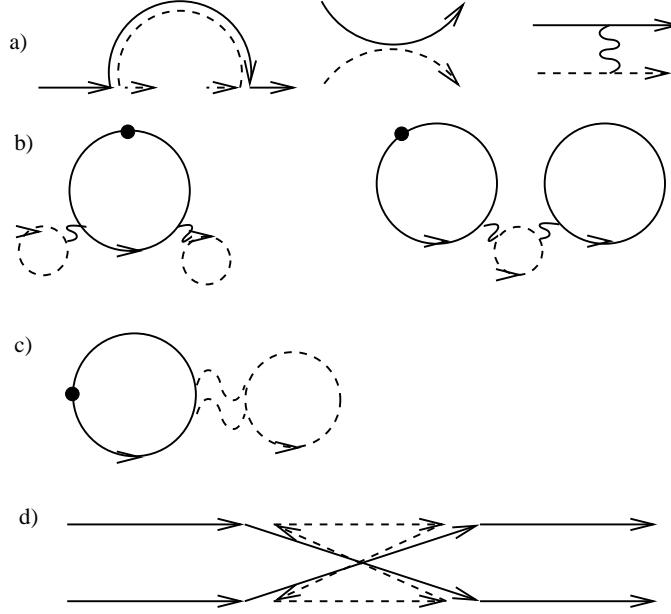

FIG. 5: (a) Deformation of diagram. (b) Deformation of diagrams in Fig. 4(a,b). (c) Example of a diagram which contributes to the annealed average but not the quenched average. (d) Same diagram as in (c).

diagrams in this restricted class, $c_{m,r}^k$, obeys

$$(m+1)k - l_n - kl_p = d > 0 \ \rightarrow \ c_{m,r}^k(l_n, l_p) \ \leq \ c_{m-d,r}^k(l_n, l_p)[m(m-1)]^{d/k} \tag{30}$$

Now, we consider a diagram which is not in this restricted class. Locate any vertex with incoming solid lines on the horizontal and an outgoing dashed line on the horizontal, such that not all pairs of lines leaving this vertex reconnect at the same vertex. Call this vertex v_1. Then, find any other vertex to which a pair of lines leaving vertex v_1 reconnects. Call this vertex v_2. Let there be l pairs of lines leaving vertex v_1 which do not conenct to v_2, and similarly l pairs of lines entering v_2 which do not come from v_1, with $1 \leq l \leq k-1$. Label these pairs of lines $L_1^1, L_2^1, ..., L_k^1$ and $L_1^2, L_2^2, ..., L_k^2$, respectively. Let these lines connect to pairs of lines $M_1^1, M_2^1, ..., M_k^1$ and $M_1^2, M_2^2, ..., M_k^2$ respectively. Let v_3 be the vertex just to the right of v_1, so that the dashed line entering v_1 comes from v_3, and similarly let v_4 be the vertex just to the left of v_2, so that the dashed line leaving v_2 goes into v_4, as shown in Fig. 6(a). Then, we determine if there is a way to re-connect pairs of lines so that now $L_{l'}^1$ connects to $L_{l'}^2$ and $M_{l'}^1$ connects to $M_{l'}^2$ for all l' in some subset of $\{1, ..., l\}$ such that the diagram splits into exactly two disconnected pieces. If there is, then we find the smallest subset of $\{1, ..., l\}$ with this property (making an arbitrary choice if there is more than one such subset) and make those reconnections. Let $\mathcal{V}_1, \mathcal{V}_2$ denote the two disconnected subsets of vertices after the reconnections. By making these reconnections, then, we are reconnecting precisely the pairs of lines which originally connected vertices in set $\mathcal{V}_1$ to those in $\mathcal{V}_2$; if there are l_c such lines, then we increase l_n by $l_c \geq 1$. Thus, we increase l_p by one and also increases l_n by at least 1. We then modify the diagram to rejoin the two pieces: the dashed line leaving to the right of vertex v_1 connects to it some vertex w_1 in the same piece, and there is some other dashed line in the other piece which connects two vertices v_1', w_1'; we re-connect these dashed lines so that v_1 connects to w_1' and v_1' connects to w_1. This reduces l_p by 1 back to its original value and makes the diagram connected. Thus, we succeed in increasing $l_n + kl_p - mk$ by at least 1.

On the other hand, if no such subset exists, we re-connect all pairs of lines for all $1 \leq l' \leq l$, as shown in Fig. 6(b). The resulting diagram must be connected (if not, then there would have been a subset of lines which could be re-connected to split the line into exactly two disconnected pieces). Then, there are two cases: the first case is when the dashed line leaving v_2 does not connect to v_1 (so that $v_2 \neq v_3$ and $v_1 \neq v_4$) *and* it is possible to re-connect the dashed lines joining v_1 to v_3 and v_2 to v_4 so that now v_2 is connected to v_1 and v_3 is connected to v_4 without breaking the diagram into two disconnected pieces. In this first case, we then also make this re-connection of dashed lines, which increases l_p by one, while keeping the diagram connected. However, in this case, the initial re-connection of pairs of lines may have reduced l_n by at most l. Thus, in this case $kl_p + l_n - mk$ is increased by at least 1. The second case is when either v_2 connects to v_1 already or it is not possible to make the re-connection of dashed lines without splitting the diagram into two pieces. This is the case in Fig. 6(b). In this case, however, l_n must have increased by at least 1
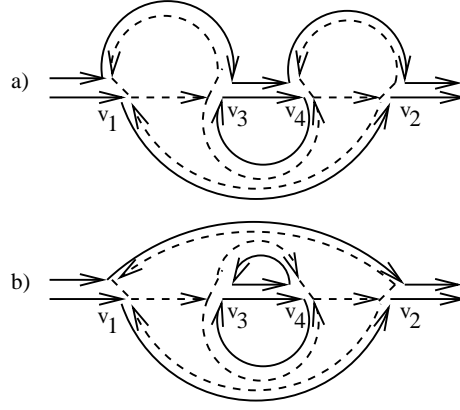
FIG. 6: (a) Diagram of Fig. 2(b) with vertices v_1, v_2, v_3, v_4 marked, for a particular choice of v_1, v_4. (b) Result of applying re-connection procedure to diagram.

by the initial re-connection of pairs of lines[3] and thus again we increase $kl_p + l_n - mk$ by at least 1.

Repeating this procedure, we ultimately arrive at a diagram in the restricted class above. At each step, we succeed in reducing $d = (m + 1)k - l_n - kl_p$ by at least unity, either by increasing l_n by at least 1 and l_p by 2, or by increasing l_p by 1 and reducing l_n by at most $k - 1$. Given a diagram in the restricted class, we can further reduce d following Eq. (30). Then, any diagram can be found by starting with a diagram in the restricted class and undoing this procedure; at each step in undoing the procedure we have at most $m^2(m-1)^{2(k-1)}$ choices (there are at most m^2 choices for v_1, v_2, and then we must re-connect at most $2(k-1)$ pairs of lines). Thus, for $(m+1)k - l_n - kl_p = d > 0$ we have

$$
\begin{aligned}
c_m^k(l_n, l_p) \ \leq \ & m^{2k} c_m^k(l_n + 1, l_p) \\
& + m^{2k} c_m^k(l_n - k + 1, l_p + 1) \\
& + m^2 c_{m-1}^k(l_n, l_p).
\end{aligned}
\tag{31}
$$

This implies that for $d > 0$,

$$
\sum_{l_n, l_p}^{l_n + l_p = m + 1 - d} c_m^k(l_n, l_p) d^{l_n} p^{l_p} \ \leq \ \delta \sum_{l_n, l_p}^{l_n + l_p = m + 1 - (d-1)} c_m^k(l_n, l_p) d^{l_n} p^{l_p},
\tag{32}
$$

with

$$
\delta = \frac{m^{2k}}{d} + \frac{m^{2k} d^{k-1}}{p} + \frac{m^2}{d^k} = (1 + x^{-1})\frac{m^{2k}}{d} + \frac{m^2}{d^k}.
\tag{33}
$$

E.   Bound on Number of Rainbow Diagrams

Finally, we provide a bound on the number of rainbow diagrams. Let us define $S_v^a(l_n, l_p)$ to equal to the number of open rainbow diagrams with solid lines at the end, with v vertices, l_n loops of solid lines (not counting the loop that would be formed by connected the open ends), and l_p disconnected sets of dashed lines, which may be constructed by at most a iterations of the process shown in Fig. 3. Similarly, define $D_v^a(l_n, l_p)$ to equal to the number of open rainbow diagrams with dashed lines at the end, with v vertices, l_n loops of solid lines (not counting the loop that would be formed by connected the open ends), and l_p disconnected sets of dashed lines, which may be constructed by at most a iterations of the process shown in Fig. 3. These open rainbow diagrams obey $l_n/k + l_p = m$. Define the

---

[3] To see why l_n must have been increased by at least one when reconnecting pairs of lines, in the case where making the reconnection of the dashed line would split the diagram into two disconnected pieces, let $\mathcal{V}_1, \mathcal{V}_2$ denote the vertices in these two disconnected pieces. Then, by reconnecting the pairs of lines, there are no longer any solid lines joining $\mathcal{V}_1$ to $\mathcal{V}_2$, so l_n increases by $l' \geq 1$.

generating function[4]

$$G_s^{(a)}(z,d,p) = \sum_v \sum_{l_n} \sum_{l_p} z^{-v/2} d^{-vk/2} d^{l_n} p^{l_p} S_v^a(l_n, l_p), \tag{34}$$

$$G_d^{(a)}(z,d,p) = \sum_v \sum_{l_n} \sum_{l_p} z^{-v/2} d^{-vk/2} d^{l_n} p^{l_p} D_v^a(l_n, l_p).$$

Then, we have the recursion relations, which come from Fig. 3(b,c):

$$G_s^{(a)}(z,d,p) = 1 + z^{-1} x G_d^{(a-1)}(z,d,p) G_s^{(a-1)}(z,d,p), \tag{35}$$

$$G_d^{(a)}(z,d,p) = 1 + z^{-1} G_s^{(a-1)}(z,d,p) G_d^{(a-1)}(z,d,p).$$

First, consider the case $x \leq 1$. From Eq. (35), $G_d^{(a)}(z,d,p) = 1 + (G_s^{(a)}(z,d,p) - 1)/x$ for all a, so that we have the recursion $G_s^{(a)}(z,d,p) = 1 + z^{-1} x G_s^{(a-1)}(z,d,p)(1 + (G_s^{(a-1)}(z,d,p) - 1)/x) = 1 + z^{-1}(x - 1) G_s^{(a-1)}(z,d,p) + z^{-1} G_s^{(a-1)}(z,d,p)^2$. The fixed points of this recursion relation are given by

$$G_s(z,d,p) \equiv \frac{z^{-1}(1-x) + 1 \pm \sqrt{(z^{-1}(1-x)+1)^2 - 4z^{-1}}}{2z^{-1}}. \tag{36}$$

Define

$$z_0 = \left(\frac{1 + x - 2\sqrt{x}}{(1-x)^2} \right)^{-1} = (1 + \sqrt{x})^2. \tag{37}$$

Then, for $z > z_0$, Eq. (35) has two real fixed points, while at $z = z_0$, Eq. (35) has a single fixed point at

$$G_s(z_0, d, p) = \frac{z_0}{2} \left[1 + z_0^{-1}(1-x) \right] = 1 + \sqrt{x} = \sqrt{z_0} > 1. \tag{38}$$

Since $G_s^{(0)}(z,d,p) = G_d^{(0)}(z,d,p) = 1$ which is smaller than the fixed point, we find that $G_s^{(a)}(z,d,p)$ increases montonically with a and remains bounded above by $G_s(z,d,p)$. All rainbow diagrams with $2m$ vertices can be found after a finite number (at most m) iterations of Fig. 3(b,c) so

$$\sum_{l_n, l_p}^{l_n + l_p = m+1} c_m^k(l_n, l_p) d^{-mk} d^{l_n} p^{l_p} \leq p z_0^m G_s(z_0, d, p). \tag{39}$$

Alternately, if $x \geq 1$, we use $G_s^{(a)}(z,d,p) = 1 + (G_d^{(a)}(z,d,p) - 1)x$, to get the recursion $G_d^{(a)}(z,d,p) = 1 + z^{-1} G_d^{(a)}(z,d,p)(1 + (G_d^{(a)}(z,d,p) - 1)x) = 1 + z^{-1}(1-x) G_d^{(a-1)}(z,d,p) + z^{-1} x G_d^{(a-1)}(z,d,p)^2$. Then, again for $z = z_0$ this recursion has a single fixed point and $G_s^{(a)}(z,d,p)$ increases monotonically with a and remains bounded by $G_s(z_0, d, p)$.

F. Sum of All Diagrams

We now bound the sum of all diagrams (26) using the bound on the sum of rainbow diagrams (39) and Eq. (32).

$$\sum_{j \geq 0} \sum_{l_n, l_p}^{l_n + l_p = m+1-j} c_m^k(l_n, l_p) d^{-mk} d^{l_n} p^{l_p} \leq p z_0^m G_s(z_0, d, p) \sum_{j \geq 0} \delta^j. \tag{40}$$

Then, if $\delta < 1$ we have

$$\hat{E}_{p,d,k}^m \leq \frac{p}{1-\delta} z_0^m G_s(z_0, d, p) = \frac{p}{1-\delta} z_0^{m + \frac{1}{2}} \tag{41}$$

---

[4] The limit as $a \to \infty$ of this generating functional is equal to, up to a factor $1/z$ in front, the Green's function in the large-d limit usually defined in field theory.

We can pick m of order $d^{1/2k}$ and still have $\delta \leq 1/2$. Then we can use $\mathbb{E}[\|\hat{M}_{p,d,k}\|] \leq (\hat{E}^m_{p,d,k})^{1/m}$ to bound

$$\mathbb{E}[\|\hat{M}_{p,d,k}\|] \leq (1 + \sqrt{x})^2 \cdot \exp\left(\frac{\ln(2p\sqrt{z_0})}{m}\right)$$

$$= (1 + \sqrt{x})^2 + O\left(\frac{k\ln(d)}{d^{\frac{1}{2k}}}\right), \tag{42}$$

as claimed in Corollary 3. We are assuming in the $O()$ notation in this bound that $x = \Theta(1)$.

III.   APPROACH 2: COMBINATORICS AND REPRESENTATION THEORY

This section gives a second proof of Theorem 1 that uses facts about symmetric subspaces along with elementary combinatorics. The fundamentals of the proof resemble those of the last section in many ways, which we will discuss at the end of this section. However, the route taken is quite different, and this approach also suggests different possible extensions.

Recall that we would like to estimate

$$E^m_{p,d,k} = \sum_{\vec{s} \in [p]^m} E_d[\vec{s}]^k.$$

Our strategy will be to repeatedly reduce the string $\vec{s}$ to simpler forms. Below we will describe two simple methods for reducing $\vec{s}$ into a possibly shorter string $R(\vec{s})$ such that $E_d[\vec{s}]$ equals $E_d[R(\vec{s})]$, up to a possible multiplicative factor of $1/d$ to some power. Next we will consider two important special cases. First are the *completely reducible* strings: $\vec{s}$ for which the reduced string $R(\vec{s})$ is the empty string. These are analogous to the rainbow diagrams in Section II and their contribution can be calculated exactly (in Section III A). The second special case is when $\vec{s}$ is *irreducible*, meaning that $R(\vec{s}) = \vec{s}$; that is, neither simplification steps can be applied to $\vec{s}$. These strings are harder to analyze, but fortunately make a smaller contribution to the final sum. In Section III B, we use representation theory to give upper bounds for $E_d[\vec{s}]$ for irreducible strings $\vec{s}$, and thereby to bound the overall contribution from irreducible strings. Finally, we can describe a general string as an irreducible string punctuated with some number of repeated letters (defined below) and completely reducible strings. The overall sum can then be bounded using a number of methods; we will choose to use a generating function approach, but inductively verifying the final answer would also be straightforward.

Reducing the string: Recall that $E_d[\vec{s}] = \text{tr}\,\varphi_{s_1} \cdots \varphi_{s_m}$, where each $|\varphi_s\rangle$ is a unit vector randomly chosen from $\mathbb{C}^d$. We will use the following two reductions to simplify $\vec{s}$.

1. *Remove repeats.* Since φ_a is a pure state, $\varphi_a^2 = \varphi_a$ and we can replace every instance of aa with a in $\vec{s}$ without changing $E_d[\vec{s}]$. Repeatedly applying this means that if $s_i = s_{i+1} = \cdots = s_j$, then $E_d[\vec{s}]$ is unchanged by deleting positions $i+1, \ldots, j$. Here we identify position i with $m + i$ for all i, so that repeats can wrap around the end of the string: e.g. the string 11332221 would become 321.

2. *Remove unique letters.* Since $\mathbb{E}[\varphi_a] = I/d$ for any a, we can replace any letters which appear only once with I/d. Thus, if $s_i \neq s_j$ for all $j \neq i$ then $E_d[\vec{s}] = E_d[\vec{s}']/d$, where $\vec{s}' \in [p]^{m-1}$ is obtained from $\vec{s}$ by deleting the position i. Repeating this process results in a string where every letter appears at least twice and with a multiplicative factor of $1/d$ for each letter that has been removed. Sometimes the resulting string will be empty, in which case we say $E_d[\emptyset] = \text{tr}\,I = d$. Thus for strings of length one, $E_d[a] = E_d[\emptyset]/d = d/d = 1$.

We will repeatedly apply these two simplification steps until no further simplifications are possible. Let $R(\vec{s})$ denote the resulting (possibly empty) string. Recall from above that when $R(\vec{s}) = \emptyset$, we say $\vec{s}$ is completely reducible, and when $R(\vec{s}) = \vec{s}$, we say $\vec{s}$ is irreducible. The sums over these two special cases are described by the following two Lemmas.

Lemma 8.

$$\frac{1}{d^k} \sum_{\substack{\vec{s} \in [p]^m \\ R(\vec{s}) = \emptyset}} E_d[\vec{s}]^k = \sum_{\ell=1}^m N(m, \ell)\frac{(p)_\ell}{d^{kl}} \leq \beta_m\left(\frac{p}{d^k}\right) \leq \lambda_+^m. \tag{43}$$

We will prove this Lemma and discuss its significance in Section III A. It will turn out that the completely reducible strings make up the dominant contribution to $E_{p,d,k}^m$ when m is not too large. Since (43) is nearly independent of k (once we fix x and p), this means that $E_{p,d,k}^m$ is also nearly independent of k. It remains only to show that the sub-leading order terms do not grow too quickly with k. Note that this Lemma establishes the lower bound of Theorem 1.

For the irreducible strings we are no longer able to give an exact expression. However, when m is sufficiently small relative to d and p, we have the following nearly tight bounds.

Lemma 9. *If* $m < \min(d^{k/6}/2^{1+k/2}, (p/5000)^{\frac{1}{2k+12}})$ *then*

$$\sum_{\substack{\vec{s}\in[p]^m \\ R(\vec{s})=\vec{s}}} E_d[\vec{s}]^k \leq \frac{e^{\frac{km}{2d^{1/3}}}}{\left(1 - \frac{5000m^{2k+12}}{p}\right)\left(1 - \frac{2^{2+k}m^2}{d^{\frac{k}{3}}}\right)} x^{\frac{m}{2}} \tag{44}$$

Additionally, when m *is even, the left-hand side of (44) is* $\geq x^{\frac{m}{2}} e^{-\frac{m^2}{2p}}$.

The proof is in Section III B. Observe that when $m \in o(d^{k/6}) \cap o(p^{\frac{1}{2k+12}})$ and m is even, we bound the sum on the LHS of (44) by $(1 \pm o(1))x^{m/2}$. We also mention that there is no factor of $1/d^k$ on the LHS, so that when $x = O(1)$ and m satisfies the above condition, the contribution from irreducible strings is a $O(1/d^k)$ fraction of the contribution from completely reducible strings.

Next, we combine the above two Lemmas to bound all strings that are not covered by Lemma 8.

Lemma 10. *If* $m < \min(d^{k/6}/2^{1+k/2}, (p/5000)^{\frac{1}{2k+12}})$ *then*

$$\sum_{\substack{\vec{s}\in[p]^m \\ R(\vec{s})\neq\emptyset}} E_d[\vec{s}]^k \leq \frac{e^{\frac{km}{2d^{1/3}}}}{\left(1 - \frac{5000m^{2k+12}}{p}\right)\left(1 - \frac{2^{2+k}m^2}{d^{\frac{k}{3}}}\right)} m\lambda_+^{m+\frac{1}{2}}. \tag{45}$$

The proof is in Section III C.

To simplify the prefactor in (45), we assume that $m < \min(2d^{1/3}/k, d^{k/6}/2^{2+k/2}, (p/5000)^{\frac{1}{2k+12}}/2)$, so that the RHS of (45) becomes simply $\leq 12m\lambda_+^{m+\frac{1}{2}}$. By Lemma 2, this is $\leq \frac{24m^3\lambda_+^2}{x}\beta_m(x)$. Then we combine Lemma 8 and Lemma 10 to obtain the bound

$$e_{p,d,k}^m \leq \left(1 + \frac{24m^3\lambda_+^2}{p}\right)\beta_m(x) \tag{46}$$

which is a variant of the upper-bound in Theorem 1. It is tighter than (7), but holds for a more restricted set of m. If we express the upper bound in terms of λ_+^m then we can skip Lemma 8 and obtain simply

$$e_{p,d,k}^m \leq \left(1 + \frac{12m\sqrt{\lambda_+}}{d^k}\right)\lambda_+^m. \tag{47}$$

A. Completely reducible strings

We begin by reviewing some facts about Narayana numbers from [30, 31]. The Narayana number

$$N(m,\ell) = \frac{1}{m}\binom{m}{\ell-1}\binom{m}{\ell} = \frac{1}{\ell}\binom{m}{\ell-1}\binom{m-1}{\ell-1} \tag{48}$$

counts the number of valid bracketings of m pairs of parentheses in which the sequence () appears ℓ times. A straightforward combinatorial proof of (48) is in [30]. When we sum (48) over ℓ (e.g. if we set $x = 1$ in (43)) then we obtain the familiar Catalan numbers $\frac{1}{m+1}\binom{2m}{m}$.

We can now prove Lemma 8. The combinatorial techniques behind the Lemma have been observed before[30, 31], and have been applied to the Wishart distribution in [10, 13].

Proof: For a string $\vec{s}$ such that $R(\vec{s}) = \emptyset$, let ℓ be the number of distinct letters in $\vec{s}$. In the process of reducing $\vec{s}$ to the empty string we will ultimately remove ℓ unique letters, so that $E_d[\vec{s}]^k = d^{k(1-\ell)}$. It remains now only to count the number of different $\vec{s}$ that sastify $R(\vec{s}) = \emptyset$ and have ℓ distinct letters.

Suppose the distinct letters in $\vec{s}$ are $S_1, S_2, \ldots, S_\ell \in [p]$. We order them so that the first occurrence of S_i is earlier than the first occurrence of S_{i+1} for each i. Let $\vec{\sigma}$ be the string obtained from $\vec{s}$ by replacing each instance of S_i with i. Then $\vec{\sigma}$ has the first occurrences of $1, 2, \ldots, \ell$ appearing in increasing order and still satisfies $R(\vec{\sigma}) = \emptyset$ and $E_d[\vec{\sigma}]^k = d^{k(1-\ell)}$. Also, for each $\vec{\sigma}$, there are $p!/(p-\ell)! \leq p^\ell$ corresponding $\vec{s}$.

It remains only to count the number of distinct $\vec{\sigma}$ for a given choice of m and ℓ. We claim that this number is given by $N(m, \ell)$. Given $\vec{\sigma}$, define a_i to be the location of the first occurrence of the letter i for $i = 1, \ldots, \ell$. Observe that

$$1 = a_1 < a_2 < \cdots < a_\ell \leq m. \tag{49}$$

Next, define μ_i to be the total number of occurrences of i in $\vec{\sigma}$, and define $b_i = \sum_{j=1}^i \mu_j$ for $i = 1, \ldots, \ell$. Then

$$1 \leq b_1 < b_2 < \cdots < b_\ell = m \tag{50}$$

Finally, we have

$$a_i \leq b_i \qquad \text{for each } i = 1, \ldots, \ell. \tag{51}$$

Ref. [31] proved that the number of $(a_1, b_1), \ldots, (a_\ell, b_\ell)$ satisfying (49), (50) and (51) is $N(m, \ell)$. Thus, we need only prove that $\vec{\sigma}$ is uniquely determined by $(a_1, b_1), \ldots, (a_\ell, b_\ell)$. The algorithm for finding $\vec{\sigma}$ is as follows.

For $t = 1, \ldots, m$.
> If $t = a_i$ then set $s := i$.
> Set $\sigma_t := s$.
> Set $\mu_s := \mu_s - 1$.
> While $(\mu_s = 0)$ set $s := s - 1$.

In other words, we start by placing 1's until we reach a_2. Then we start placing 2's until we've either placed μ_2 2's, in which case we go back to placing 1's; or we've reached a_3, in which we case we start placing 3's. The general rule is that we keep placing the same letter until we either encounter the next a_i or we run out of the letter we were using, in which case we go back to the last letter we placed.

To show that $\vec{\sigma}$ couldn't be constructed in any other way, first note that we have $\sigma_{a_i} = i$ for each i by definition. Now fix an i and examine the interval between a_i and a_{i+1}. Since it is before a_{i+1}, it must contain only letters in $\{1, \ldots, i\}$. Using the fact that $R(\vec{\sigma}) = \emptyset$, we know that $\vec{\sigma}$ cannot contain the subsequence j-i-j-i (i.e. cannot be of the form $\cdots j \cdots i \cdots j \cdots i$). We now consider two cases.

Case (1) is that $\mu_i \geq a_{i+1} - a_i$. In this case we must have $\sigma_t = i$ whenever $a_i < t < a_{i+1}$. Otherwise, this would mean that some $s \in \{1, \ldots, i-1\}$ appears in this interval, and since s must have appeared earlier as well ($s < i$ so $a_s < a_i$ and $\sigma_{a_s} = s$), then no i's can appear later in the string. However, this contradicts the fact that $\mu_i \geq a_{i+1} - a_i$. Thus if $\mu_i \geq a_{i+1} - a_i$ then the entire interval between a_i and a_{i+1} must contain i's.

Case (2) is that $\mu_i < a_{i+1} - a_i$. This means that there exists t with $a_i < t < a_{i+1}$ and $\sigma_t \in \{1, \ldots, i-1\}$; if there is more than one then take t to be the lowest (i.e. earliest). Note that $\sigma_{t'} \neq i$ for all $t' > t$; otherwise we would have a σ_t-i-σ_t-i subsequence. Also, by definition $\sigma_{t'} = i$ for $a_i \leq t' < t$. Since this is the only place where i appears in the string, we must have $t = a_i + \mu_i$. Once we have placed all of the i's, we can proceed inductively to fill the rest of the interval with letters from $\{1, \ldots, i-1\}$.

In both cases, $\vec{\sigma}$ is uniquely determined by $a_1, \ldots, a_\ell$ and $b_1, \ldots, b_\ell$ (or equivalently, $\mu_1, \ldots, \mu_\ell$). This completes the proof of the equality in (43). $\square$

Before continuing, we will mention some facts about Narayana numbers that will later be useful. Like the Catalan numbers, the Narayana numbers have a simple generating function; however, since they have two parameters the generating function has two variables. If we define

$$F(x, y) = \sum_{0 \leq \ell \leq m < \infty} N(m, \ell) x^\ell y^m, \tag{52}$$

then one can show[30, 31] (but note that [30] takes the sum over $m \geq 1$) that

$$F(x, y) = \frac{1 + (1-x)y - \sqrt{1 - 2(1+x)y + (1-x)^2 y^2}}{2y}. \tag{53}$$

We include a proof for convenience. First, by convention $N(0, 0) = 1$. Next, an arrangement of m pairs of parentheses can start either with () or ((. Starting with () leaves $N(m-1, \ell-1)$ ways to complete the string. If the string starts

with *((* then suppose the *)* paired with the first *(* is the ith *)* in the string. We know that $2 \le i \le m$ and that the first $2i$ characters must contain exactly i *(*'s and i *)*'s. Additionally, the $2i - 1^{\text{st}}$ and $2i^{\text{th}}$ characters must both be *)*'s. Let j be the number of appearances of *()* amongst these first $2i$ characters. Note that $j \le \min(i-1, \ell)$, and that *()* appears $\ell - i$ times in the last $2m - 2i$ characters. Thus there are

$$\sum_{i=2}^{m} \sum_{j=1}^{\min(i-1,\ell)} N(i-1,j)N(m-i,\ell-j) = -N(m-1,\ell) + \sum_{i=1}^{m} \sum_{j=1}^{\min(i-1,\ell)} N(i-1,j)N(m-i,\ell-j)$$

ways to complete a string starting with *((*. Together, these imply that

$$N(m,\ell) = N(m-1,\ell-1) - N(m-1,\ell) + \sum_{i=1}^{m} \sum_{j=1}^{\min(i-1,\ell)} N(i-1,j)N(m-i,\ell-j), \tag{54}$$

which we can state equivalently as an identify for the generating function (52):

$$F = 1 + xyF + y(F^2 - F), \tag{55}$$

which has the solution (53). (The sign in front of the square root can be established from $1 = N(0,0) = F(x,0)$.)

Connection to Section II: Observe that (53) matches (36) once we make the substitution $y = z^{-1}$. Indeed it can be shown that rainbow diagrams have a one-to-one correspondence with valid arrangements of parentheses, and thus can be enumerated by the Narayana numbers in the same way.

Connection to free probability: Another set counted by the Narayana numbers is the set of noncrossing partitions of $[m]$ into ℓ parts. The non-crossing condition means that we never have $a < b < c < d$ with a, c in one part of the partition and b, d in another; it is directly analogous to the property that $\vec{\sigma}$ contains no subsequence of the form j-i-j-i.

To appreciate the significance of this, we return to the classical problem of throwing p balls into d bins. The occupancy of a single bin is $z = z_1 + \ldots + z_p$ where $z_1, \ldots, z_p$ are i.i.d. and have $\Pr[z_i = 0] = 1 - 1/d$, $\Pr[z_i = 1] = 1/d$. One can readily verify that

$$\mathbb{E}[z^m] = \sum_{\ell=1}^{m} |\operatorname{Par}(m,\ell)| \frac{(p)_\ell}{d^\ell},$$

where $\operatorname{Par}(m,\ell)$ is the set of (unrestricted) partitions of m into ℓ parts.

This is an example of a more general phenomenon in which convolution of classical random variables involves partitions the same way that convolution of free random variables involves non-crossing partitions. See Ref. [29] for more details.

B. Irreducible strings

As with the completely reducible strings, we will break up the sum based on the powers of p and d which appear. However, while in the last section p and $1/d^k$ both depended on the single parameter ℓ, here we will find that some terms are smaller by powers of $1/p$ and/or $1/d$. Our strategy will be to identify three parameters—ℓ, c_2, and $\hat{\mu}_2$, all defined below—for which the leading contribution occurs when all three equal $m/2$. We show that this contribution is proportional to $\sqrt{x}^m$ and that all other values of ℓ, c_2, and $\hat{\mu}_2$ make negligible contributions whenever m is sufficiently small.

Again, we will let ℓ denote the number of unique letters in $\vec{s}$. We will also let $S_1, \ldots, S_\ell \in [p]$ denote these unique letters. However, we choose them so that $S_1 < S_2 < \cdots < S_\ell$, which can be done in

$$\binom{p}{\ell} \le \frac{p^\ell}{\ell!} \tag{56}$$

ways. Again, we let $\vec{\sigma} \in [\ell]^m$ be the string that results from replacing all the instances of S_i in $\vec{s}$ with i. However, because of our different choice of $S_1, \ldots, S_\ell$, we no longer guarantee anything about the ordering of $1, \ldots, \ell$ in $\vec{\sigma}$.

We will also take μ_a be the frequency of a in $\vec{\sigma}$ for each $a = 1, \ldots, \ell$. We also define $\hat{\mu}_b$ to be the number of a such that $\mu_a = b$. Observe that

$$\ell = \sum_b \hat{\mu}_b \tag{57}$$

$$m = \sum_{a=1}^{\ell} \mu_a = \sum_b b \hat{\mu}_b. \tag{58}$$

Also recall that since $R(\vec{\sigma}) = \vec{\sigma}$, $\vec{\sigma}$ has no repeats or unique letters. Thus $\mu_a \geq 2$ for each a, or equivalently $\hat{\mu}_1 = 0$. This also implies that $\ell \leq m/2$. Since (56) is maximised when $\ell = \lfloor \frac{m}{2} \rfloor$, we will focus on this case first and then show that other values of ℓ have smaller contributions. Moreover (58) implies that $\hat{\mu}_2 \leq \ell$ and (57), (58) and the fact that $\hat{\mu}_1 = 0$ imply that $\hat{\mu}_2 \geq 3\ell - m$. Together we have

$$3\ell - m \leq \hat{\mu}_2 \leq \ell. \tag{59}$$

Thus ℓ is close to $m/2$ if and only if $\hat{\mu}_2$ is as well. This will be useful because strings will be easier to analyze when almost all letters occur exactly twice.

We now turn to the estimation of $E_d[\vec{\sigma}]$. To analyze $E_d[\vec{\sigma}] = \mathbb{E}[\operatorname{tr} \varphi_{\sigma_1} \varphi_{\sigma_2} \cdots \varphi_{\sigma_m}]$, we first introduce the cyclic shift operator

$$C_m = \sum_{i_1, \ldots, i_m \in [d]} |i_1, \ldots, i_m\rangle \langle i_2, \ldots, i_m, i_1|.$$

Then we use the identity

$$\operatorname{tr}[\varphi_{\sigma_1} \varphi_{\sigma_2} \cdots \varphi_{\sigma_m}] = \operatorname{tr}[C_m(\varphi_{\sigma_1} \otimes \varphi_{\sigma_2} \otimes \cdots \otimes \varphi_{\sigma_m})]. \tag{60}$$

Next, we take the expectation. It is a well-known consequence of Schur-Weyl duality (see e.g. Lemma 1.7 of [11]) that

$$\mathbb{E}[\varphi^{\otimes t}] = \frac{\sum_{\pi \in \mathcal{S}_t} \pi}{d(d+1) \cdots (d+t-1)}. \tag{61}$$

We will apply this to (60) by inserting (61) in the appropriate locations as given by $\vec{\sigma}$. Let $\mathcal{S}_{\vec{\sigma}} := \{\pi \in \mathcal{S}_m : \sigma_i = \sigma_{\pi(i)} \forall i \in [m]\}$ be the set of permutations that leaves $\vec{\sigma}$ (or equivalently $\vec{s}$) invariant. Then $|\mathcal{S}_{\vec{\sigma}}| = \mu_1! \cdots \mu_\ell!$ and

$$E_d[\vec{\sigma}] = \mathbb{E}[\operatorname{tr} C_m(\varphi_{\sigma_1} \otimes \varphi_{\sigma_2} \otimes \cdots \otimes \varphi_{\sigma_m})] \tag{62a}$$

$$= \operatorname{tr} C_m \frac{\sum_{\pi \in \mathcal{S}_{\vec{\sigma}}} \pi}{\prod_{i=1}^{\ell} d(d+1) \cdots (d+\mu_i-1)} \tag{62b}$$

$$\leq \operatorname{tr} C_m \frac{\sum_{\pi \in \mathcal{S}_{\vec{\sigma}}} \pi}{\prod_{i=1}^{\ell} d^{\mu_i}} \tag{62c}$$

$$= \frac{\sum_{\pi \in \mathcal{S}_{\vec{\sigma}}} \operatorname{tr} C_m \pi}{d^m} \tag{62d}$$

$$= \sum_{\pi \in \mathcal{S}_{\vec{\sigma}}} d^{\operatorname{cyc}(C_m \pi) - m}. \tag{62e}$$

This last equality follows from the fact that for any permutation ν acting on $(\mathbb{C}^d)^{\otimes m}$, we have that

$$\operatorname{tr} \nu = d^{\operatorname{cyc}(\nu)}, \tag{63}$$

where $\operatorname{cyc}(\nu)$ is the number of cycles of ν. (Eq. (63) can be proven by first considering the case when $\operatorname{cyc}(\nu) = 1$ and then decomposing a general permutation into a tensor product of cyclic permutations.)

To study $\operatorname{cyc}(C_m \pi)$, we introduce a graphical notation for strings. For any string $\vec{\sigma}$, define the *letter graph* G to be a directed graph with ℓ vertices such that for $i = 1, \ldots, \ell$, vertex i has in-degree and out-degree both equal to μ_i. (For brevity, we will simply say that i has degree μ_i.) Thus there are a total of m edges. The edges leaving and entering vertex i will also be ordered. To construct the edges in G, we add an edge from s_i to s_{i+1} for $i = 1, \ldots, m$, with $s_{m+1} := s_1$. The ordering on these edges is given by the order we add them in. That is, if letter a appears in
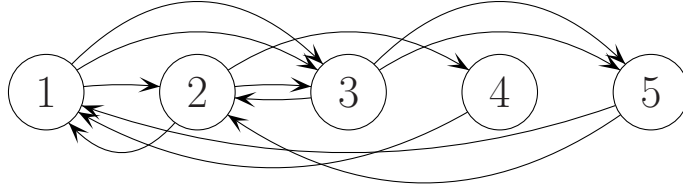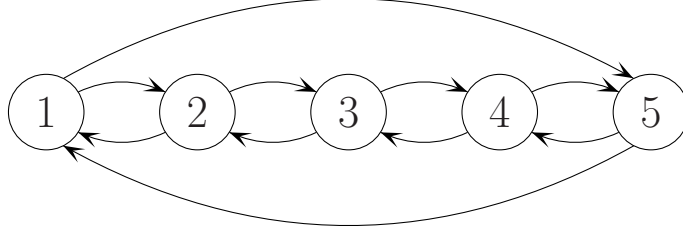
FIG. 7: The letter graph corresponding to the string 123241351352.

FIG. 8: Example of the letter graph for the case when $m = 10$ and $\ell = \hat{\mu}_2 = c_2^{\max} = \frac{m}{2}$. The corresponding string is 1234543215.

positions $i_1, i_2, \ldots$ with $i_1 < i_2 < \cdots$, then the first edge out of a is directed at s_{i_1+1}, the second out-edge points at s_{i_2+1}, and so on. Likewise, a's incoming edges (in order) come from $s_{i_1-1}, s_{i_2-1}, \ldots$.

Now we think of the incoming and outgoing edges of a vertex as linked, so that if we enter on the jth incoming edge of a vertex, we also exit on the jth outgoing edge. This immediately specifies a cycle through some or all of G. If we use the ordering specified in the last paragraph then the cycle is in fact an Eulerian cycle (i.e. visits each edge exactly once) that visits the vertices in the order $s_1, s_2, \ldots, s_m$. Thus, from a letter graph G and a starting vertex we can reconstruct the string $\vec{\sigma}$ that was used to generate G.

The letter graph of $\vec{\sigma}$ can also be used to give a cycle decomposition of $C_m\pi$. Any permutation $\pi \in \mathcal{S}_{\vec{\sigma}}$ can be thought of as permuting the mapping between in-edges and out-edges for each vertex. The resulting number of edge-disjoint cycles is exactly $\text{cyc}(C_m\pi)$. To see this, observe that π maps i_1 to some i_2 for which $\sigma_{i_1} = \sigma_{i_2}$ and then C_m maps i_2 to $i_2 + 1$. In G these two steps simply correspond to following one of the edges out of i_1. Following the path (or the permutation) until it repeats itself, we see that cycles in G are equivalent to cycles in $C_m\pi$.

We now use letter graphs to estimate (62). While methods for exactly enumerating cycle decompositions of directed graphs do exist[7], for our purposes a crude upper bound will suffice. Observe that because $\vec{\sigma}$ contains no repeats, G contains no 1-cycles. Thus, the shortest cycles in G (or equivalently, in $C_m\pi$) have length 2. Let $c_2(\pi)$ denote the number of 2-cycles in $C_m\pi$ and $c_2^{\max} = \max_{\pi \in \mathcal{S}_{\vec{\sigma}}} c_2(\pi)$. Sometimes we simply write c_2 instead of $c_2(\pi)$ when the argument is understood from context. We now observe that c_2 obeys bounds analogous to those in (59). In particular, $c_2^{\max} \leq \frac{m}{2}$, and for any π,

$$\text{cyc}(C_m\pi) \leq c_2(\pi) + \frac{m - 2c_2(\pi)}{3} = \frac{m + c_2(\pi)}{3}. \tag{64}$$

Since $c_2(\pi) \leq c_2^{\max} \leq m/2$, (64) implies that $\text{cyc}(C_m\pi) \leq m/2$. Thus the smallest power of $1/d$ possible in (62) is $\frac{m}{2}$. When we combine this with (59), we see that the leading-order contribution (in terms of p and d) is $O(x^{m/2})$, and that other terms are smaller by powers of $1/p$ and/or $1/d$. Additionally, this leading-order contribution will have a particularly simple combinatorial factor.

The leading-order term. Consider the case when m is even and $\ell = \hat{\mu}_2 = c_2^{\max} = \frac{m}{2}$. This corresponds to a graph with ℓ vertices, each with in-degree and out-degree two. Additionally, there is an ordering of the edges which organizes them into ℓ 2-cycles. Thus every vertex participates in exactly two 2-cycles. Since the graph is connected, it must take the form of a single doubly-linked loop. Thus the letter graph of the leading-order term is essentially unique. See Fig. 8 for an example when $m = 10$. The only freedom here is the ordering of the vertices, which can be performed in $\ell!$ ways. Together with (56), this means the combinatorial contribution is simply $\ell!\binom{p}{\ell} \leq p^\ell$.

Now we examine the sum in (62). Assume without loss of generality that the vertices $1, \ldots, \ell$ are connected in the cycle $1 - 2 - 3 - \cdots - \ell - 1$. Each vertex has two different configurations corresponding to the two permutations in $\mathcal{S}_2$. In terms of the letter graph these correspond to the two different ways that the two incoming edges can be connected to the two outgoing edges. Since vertex i has one edge both to and from each of $i \pm 1$, we can either

FIG. 9: Vertex i is connected to $i \pm 1$ by one edge in either direction. These edges can be connected to each other in two ways, which are depicted in (a) and (b). We call (a) a "closed" configuration and (b) an "open" configuration.

(a) connect the incoming $i - 1$ edge to the outgoing $i - 1$ edge, and the incoming $i + 1$ edge to the outgoing $i + 1$ edge (the closed configuration) ; or,

(b) connect the incoming $i - 1$ edge to the outgoing $i + 1$ edge, and the incoming $i + 1$ edge to the outgoing $i - 1$ edge (the open configuration).

These possibilities are depicted in Fig. 9.

Let c denote the number of vertices in closed configurations. These vertices can be selected in $\binom{\ell}{c}$ ways. If $1 \le c \le \ell$ then c is also the number of cycles: to see this, note that each closed configuration caps two cycles and each cycle consists of a chain of open configurations that is capped by two closed configurations on either end. The exception is when $c = 0$. In this case, there are two cycles, each passing through each vertex exactly once. Thus, the RHS of (62) evaluates (exactly) to

$$d^{2-m} + \sum_{c=1}^{\ell} \binom{\ell}{c} d^{c-m} = d^{-\frac{m}{2}} \left[\left(1 + \frac{1}{d} \right)^{\frac{m}{2}} + d^{-\frac{m}{2}}(d^2 - 1) \right].$$

Combining everything, we find a contribution of $x^{\frac{m}{2}}(1 + o(1))$ as $d \to \infty$. In particular, when m is even this yields the lower bound claimed in Lemma 9. We now turn to the case when $c_2^{\max}$, ℓ and $\hat{\mu}_2$ are not all equal to $m/2$.

The sum over all terms. Our method for handling arbitrary values of $c_2^{\max}$, ℓ and $\hat{\mu}_2$ is to compare their contribution with the leading-order term. We find that if one of these variables is decreased we gain combinatorial factors, but also need to multiply by a power of $1/p$ or $1/d$. The combinatorial factors will turn out to be polynomial in m, so if m is sufficiently small the contributions will be upper-bounded by a geometrically decreasing series. This process resembles (in spirit, if not in details) the process leading to Eq. (31) in Section II.

Our strategy is to decompose the graph into a "standard" component which resembles the leading-order terms and a "non-standard" component that can be organized arbitrarily. The standard component is defined to be the set of 2-cycles between degree-2 vertices. When $\ell = \hat{\mu}_2 = c_2^{\max} = \frac{m}{2}$ the entire graph is in the standard component, so when $\ell, \hat{\mu}_2, c_2^{\max} \approx \frac{m}{2}$, the non-standard component should be small. Thus, in what follows, it will be helpful to keep in mind that the largest contributions come from when $\frac{m}{2} - \ell$, $\frac{m}{2} - \hat{\mu}_2$, $\frac{m}{2} - c_2^{\max}$ are all small, and so our analysis will focus on this case.

Begin by observing that there are $\ell - \hat{\mu}_2$ vertices with degree greater than two. Together these vertices have $m - 2\hat{\mu}_2$ in- and out-edges. Thus, they (possibly together with some of the degree-2 vertices) can participate in at most $m - 2\hat{\mu}_2$ 2-cycles. Fix a permutation π for which $c_2(\pi) = c_2^{\max}$. To account for all the 2-cycles, there must be at least $c_2^{\max} - (m - 2\hat{\mu}_2)$ 2-cycles between degree-2 vertices. These 2-cycles amongst degree-2 vertices (the standard component) account for $\ge 2c_2^{\max} - 2m + 4\hat{\mu}_2$ edges. Thus the number of non-standard edges entering and leaving the degree-2 vertices is $\le 2\hat{\mu}_2 - (2c_2^{\max} - 2m + 4\hat{\mu}_2) = 2m - 2c_2^{\max} - 2\hat{\mu}_2$. Together we have $\le 3m - 2c_2^{\max} - 4\hat{\mu}_2$ non-standard edges.

We now bound the number of ways to place the m edges in G. First, we can order the degree-2 vertices in $\hat{\mu}_2!$ ways. This ordering will later be used to place the 2-cycles of the standard component. Next, we fix an arbitrary ordering for the $\ell - \hat{\mu}_2$ vertices with degree larger than two. We then place

$$e_{\mathrm{NS}} := 3m - 2c_2^{\max} - 4\hat{\mu}_2$$

non-standard edges. This can be done in $\le m^{e_{\mathrm{NS}}}$ ways. One way to see this is that each non-standard edge has m choices of destination, since we allow them to target specific incoming edges of their destination vertex. Call these destination edges $\{I_1, \ldots, I_{e_{\mathrm{NS}}}\}$. These incoming edges correspond to e_{NS} outgoing edges, which we call $\{O_1, \ldots, O_{e_{\mathrm{NS}}}\}$, and which become the starting points of the non-standard edges. Without loss of generality we can sort $\{O_1, \ldots, O_{e_{\mathrm{NS}}}\}$ according to some canonical ordering; let $\{O'_1, \ldots, O'_{e_{\mathrm{NS}}}\}$ be the sorted version of the list. Then we let O'_i connect to I_i for $i = 1, \ldots, e_{\mathrm{NS}}$. Since our ordering of $\{I_1, \ldots, I_{e_{\mathrm{NS}}}\}$ was arbitrary, this is enough to specify

any valid placement of the edges. Additionally, our choices of $\{I_1, \ldots, I_{e_{\text{NS}}}\}$ also determine the degrees $\mu_1, \ldots, \mu_\ell$ since they account for all of the incoming edges of the non-degree-2 vertices and out-degree equals in-degree. Note that nothing prevents non-standard edges from being used to create 2-cycles between degree-2 vertices. However we conservatively still consider such cycles to be part of the non-standard component.

The remaining $m - e_{\text{NS}} = 2c_2^{\max} + 4\hat{\mu}_2 - 2m$ edges (if this number is positive) make up 2-cycles between degree-2 vertices, i.e. the standard component. Here we use the ordering of the degree-2 vertices. After the non-standard edges are placed, some degree-2 vertices will have all of their edges filled, some will have one incoming and one outgoing edge filled, and some will have none of their edges filled. Our method of placing 2-cycles is simply to place them between all pairs of neighbors (relative to our chosen ordering) whenever this is possible.

We conclude that the total number of graphs is

$$\leq \hat{\mu}_2! m^{e_{\text{NS}}} \leq \hat{\mu}_2! m^{3m-2c_2^{\max}-4\hat{\mu}_2} \leq \ell! m^{3m-2c_2^{\max}-4\hat{\mu}_2}. \tag{65}$$

In order to specify a string $\vec{\sigma}$, we need to additionally choose a starting edge. However, if we start within the standard component, the fact that we have already ordered the degree-2 vertices means that this choice is already accounted for. Thus we need only consider

$$e_{\text{NS}} + 1 \leq 2^{e_{\text{NS}}} = 2^{3m-2c_2^{\max}-4\hat{\mu}_2} \tag{66}$$

initial edges, where we have used the fact that $1 + a \leq 2^a$ for any integer a. The total number of strings corresponding to given values of $\ell, \hat{\mu}_2, c_2^{\max}$ is then upper-bounded by the product of (66) and (65):

$$\ell! (2m)^{3m-2c_2^{\max}-4\hat{\mu}_2}. \tag{67}$$

Observe that this matches the combinatorial factor for the leading-order term ($c_2^{\max} = \hat{\mu}_2 = \ell = m/2$) and then degrades smoothly as $c_2^{\max}, \hat{\mu}_2, \ell$ move away from $m/2$.

Finally, we need to evaluate the sum over permutations in (62). Our choices for non-standard vertices are substantially more complicated than the open or closed options we had for the leading-order case. Fortunately, it suffices to analyze only whether each 2-cycle is present or absent. Since a 2-cycle consists of a pair of edges of the form (i, j) and (j, i), each such cycle can independently be present or absent. Thus, while there are $\mu_1! \cdots \mu_\ell!$ total elements of $\mathcal{S}_{\vec{\sigma}}$, we can break the sum into $2^{c_2^{\max}}$ different groups of $(\mu_1! \cdots \mu_\ell!)/2^{c_2^{\max}}$ permutations, each corresponding to a different subset of present 2-cycles. In other words, there are exactly

$$\binom{c_2^{\max}}{c} \frac{\mu_1! \cdots \mu_\ell!}{2^{c_2^{\max}}}$$

choices of $\pi \in \mathcal{S}_{\vec{\sigma}}$ such that $c_2(\pi) = c$. Using the fact that $\text{cyc}(C_m \pi) \leq (m + c_2(\pi))/3$, we have

$$E_d[\vec{\sigma}] \leq \sum_{c=0}^{c_2^{\max}} \binom{c_2^{\max}}{c} \frac{\mu_1! \cdots \mu_\ell!}{2^{c_2^{\max}}} d^{\frac{m+c}{3}-m} = \frac{\mu_1! \cdots \mu_\ell!}{2^{c_2^{\max}}} d^{\frac{-2m+c_2^{\max}}{3}} \left(1 + d^{-\frac{1}{3}}\right)^{c_2^{\max}}$$

Finally, observe that $\mu_1! \cdots \mu_\ell!$ is a convex function of $\mu_1, \ldots, \mu_\ell$ and thus is maximized when $\mu_1 = m - 2\ell + 2$ and $\mu_2 = \cdots = \mu_\ell = 2$ (ignoring the fact that we have already fixed $\hat{\mu}_2$). Thus

$$E_d[\vec{\sigma}] \leq (m - 2\ell + 2)! 2^{\ell-1-c_2^{\max}} d^{\frac{-2m+c_2^{\max}}{3}} \left(1 + d^{-\frac{1}{3}}\right)^{c_2^{\max}} \tag{68}$$

$$\leq m^{m-2\ell} 2^{\frac{m}{2}-c_2^{\max}} d^{\frac{-2m+c_2^{\max}}{3}} e^{\frac{m}{2d^{1/3}}}, \tag{69}$$

where in the last step we used the facts that $2 \leq \ell \leq m/2$ and $c_2^{\max} \leq m/2$.

We now combine (69) with the combinatorial factor in (67) to obtain

$$\sum_{\substack{\vec{s} \in [p]^m \\ R(\vec{s}) = \vec{s}}} E_d[\vec{s}]^k \leq \sum_{0 \leq c_2^{\max} \leq \frac{m}{2}} \sum_{0 \leq \ell \leq \frac{m}{2}} \sum_{3\ell-m \leq \hat{\mu}_2 \leq \ell} \frac{p^\ell}{\ell!} \ell! (2m)^{3m-2c_2^{\max}-4\hat{\mu}_2} \left[m^{m-2\ell} 2^{\frac{m}{2}-c_2^{\max}} d^{\frac{-2m+c_2^{\max}}{3}} e^{\frac{m}{2d^{1/3}}} \right]^k \tag{70}$$

$$= x^{\frac{m}{2}} e^{\frac{km}{2d^{1/3}}} \sum_{\substack{0 \leq c_2^{\max} \leq \frac{m}{2} \\ 0 \leq \ell \leq \frac{m}{2} \\ 3\ell-m \leq \hat{\mu}_2 \leq \ell}} \frac{p^{\ell-\frac{m}{2}}}{d^{\frac{k}{3}\left(\frac{m}{2}-c_2^{\max}\right)}} (2m)^{(m-2c_2^{\max})+(2m-4\hat{\mu}_2)} m^{k(m-2\ell)} 2^{k\left(\frac{m}{2}-c_2^{\max}\right)} \tag{71}$$

We can bound the sum over $\hat{\mu}_2$ by introducing $\alpha = \ell - \hat{\mu}_2$, so that

$$\sum_{3\ell - m \leq \hat{\mu}_2 \leq \ell} (2m)^{2(m-2\hat{\mu}_2)} = (2m)^{2m-4\ell} \sum_{\alpha=0}^{m-2\ell} (2m)^{4\alpha} = (2m)^{2m-4\ell}(1 + 16m^4)^{m-2\ell} \leq \left(65m^6\right)^{m-2\ell} \tag{72}$$

Substituting (72) in (71) and rearranging, we obtain

$$\sum_{\substack{\vec{s} \in [p]^m \\ R(\vec{s}) = \vec{s}}} E_d[\vec{s}]^k \leq x^{\frac{m}{2}} e^{\frac{km}{2d^{1/3}}} \sum_{0 \leq c_2^{\max} \leq \frac{m}{2}} \sum_{0 \leq \ell \leq \frac{m}{2}} \left(\frac{5000m^{2k+12}}{p}\right)^{\frac{m}{2} - \ell} \left(\frac{2^{2+k}m^2}{d^{\frac{k}{3}}}\right)^{\frac{m}{2} - c_2^{\max}} \tag{73}$$

$$\leq \frac{e^{\frac{km}{2d^{1/3}}}}{\left(1 - \frac{5000m^{2k+12}}{p}\right)\left(1 - \frac{2^{2+k}m^2}{d^{\frac{k}{3}}}\right)} x^{\frac{m}{2}}. \tag{74}$$

In the last step we have assumed that both terms in the denominator are positive. This completes the proof of Lemma 9. □

C. Bounding the sum of all strings

For any string $\vec{s} \in [p]^m$ we will repeatedly remove repeats and unique letters until the remaining string is irreducible. Each letter in the original string either (a) appears in the final irreducible string, (b) is removed as a repeat of one of the letters appearing in the final irreducible string, or (c) is removed as part of a completely reducible substring. Call the letters A, B or C accordingly. Assign a weight of $\sqrt{x}^t y^t$ to each run of t A's, a weight of y^t to each run of t B's and of $\sum_{t=0}^{\infty} \sum_{\ell=0}^{t} N(t, \ell) x^\ell y^t$ to each run of t C's. Here y is an indeterminant, but we will see below that it can also be thought of as a small number. We will define $G(x, y)$ to be the sum over all finite strings of A's, B's and C's, weighted according to the above scheme. Note that $[y^m]G(x, y)$ (i.e. the coefficient of y^m in $G(x, y)$) is the contribution from strings of length m.

We now relate $G(x, y)$ to the sum in (45). Define

$$A_0 = \frac{e^{\frac{km}{2d^{1/3}}}}{\left(1 - \frac{5000m^{2k+12}}{p}\right)\left(1 - \frac{2^{2+k}m^2}{d^{\frac{k}{3}}}\right)}$$

so that Lemma 9 implies that the contribution from all irreducible strings of length t is $\leq A_0 \sqrt{x}^t$ as long as $1 \leq t \leq m$. We will treat the $t = 0$ case separately in Lemma 8, but for simplicity allow it to contribute a $A_0 \sqrt{x}^0$ term to the present sum. Similarly, we ignore the fact that there are no irreducible strings of length 1, 2, 3 or 5, since we are only concerned with establishing an upper bound here. Thus

$$\sum_{\substack{\vec{s} \in [p]^m \\ R(\vec{s}) \neq \emptyset}} E_d[\vec{s}]^k \leq A_0 [y^m] G(x, y) \leq A_0 \frac{G(x, y_0)}{y_0^m}, \tag{75}$$

where the second inequality holds for any y_0 within the radius of convergence of G. We will choose y_0 below, but first give a derivation of $G(x, y)$.

To properly count the contributions from completely reducible substrings (a.k.a. C's), we recall that $F(x, y)$ counts all C strings of length ≥ 0. Thus, it will be convenient to model a general string as starting with a run of 0 or more C's, followed by one or more steps, each of which places either an A or a B, and then a run of 0 or more C's. (We omit the case where the string consists entirely of C's, since this corresponds to completely reducible strings.) Thus,

$$G(x, y) = F(x, y) \cdot \sum_{n \geq 1} \left[y(1 + \sqrt{x})F(x, y)\right]^n = \frac{y(1 + \sqrt{x})F^2(x, y)}{1 - y(1 + \sqrt{x})F(x, y)}, \tag{76}$$

which converges whenever F converges and $y(1 + \sqrt{x})F < 1$. However, since we are only interested in the coefficient of y^m we can simplify our calculations by summing over only $n \leq m$. We also omit the $n = 0$ term, which corresponds to the case of completely reducible strings, which we treat separately. Thus, we have

$$G_m(x, y) := F(x, y) \cdot \sum_{n=1}^{m} \left[y(1 + \sqrt{x})F(x, y)\right]^n,$$

and $G_m(x,y)$ satisfies $[y^m]G_m(x,y) = [y^m]G(x,y)$.

Now define $y_0 = \lambda_+^{-1} = (1+\sqrt{x})^{-2}$. Rewriting F as $\frac{1}{2}\left(y^{-1} + 1 - x - \sqrt{(y^{-1} - (1+x))^2 - 4x}\right)$, we see that $F(x,y_0) = 1 + \sqrt{x}$. Thus $y_0(1+\sqrt{x})F(x,y_0) = 1$ and $G_m(x,y_0) = m(1+\sqrt{x})$.

Substituting into (75) completes the proof of the Lemma.

D. Alternate models

We now use the formalism from Section III B to analyze some closely related random matrix ensembles that have been suggested by the information locking proposals of [23]. The first ensemble we consider is one in which each $|\varphi_s^j\rangle$ is a random unit vector in $A_j \otimes B_j$, then the B_j system is traced out. Let $d_A = \dim A_1 = \ldots = \dim A_k$ and $d_B = \dim B_1 = \ldots = \dim B_k$. The resulting matrix is

$$M_{p,d_A[d_B],k} := \sum_{\vec{s}\in[p]^m} \bigotimes_{j=1}^k \mathrm{tr}_{B_j}\, \varphi_{s_j}^j.$$

If $d_B \ll d_A$ then we expect the states $\mathrm{tr}_B \varphi_s$ to be nearly proportional to mutually orthogonal rank-d_B projectors and so we expect $M_{p,d_A[d_B],k}$ to be nearly isospectral to $M_{p,d_A/d_B,k} \otimes \tau_{d_B}^{\otimes k}$, where $\tau_d := I_d/d$. Indeed, if we define $E_{p,d_A[d_B],k}^m := \mathrm{tr}\, M_{p,d_A[d_B],k}^m$ then we have

Lemma 11.

$$E_{p,d_A[d_B],k}^m \leq E_{p,d_A/d_B,k}^m e^{\frac{m(m+1)kd_B}{2d_A}} d_B^{k(1-m)}.$$

Proof. Define $E_{d_A[d_B]}[\vec{s}] = \mathrm{tr}(\mathrm{tr}_{B_1}(\varphi_{s_1}^1)\cdots\mathrm{tr}_{B_m}(\varphi_{s_m}^1))$. Following the steps of (62), we see that

$$E_{d_A[d_B]}[\vec{s}] = \mathrm{tr}(C_m^{A^m} \otimes I^{B^m})\mathbb{E}(\varphi_{s_1}\otimes\cdots\varphi_{s_m}) \tag{77}$$

$$\leq \mathrm{tr}(C_m^{A^m} \otimes I^{B^m})\frac{\sum_{\pi\in\mathcal{S}_{\vec{s}}} \pi^{A^m} \otimes \pi^{B_m}}{(d_A d_B)^m} \tag{78}$$

$$= \sum_{\pi\in\mathcal{S}_{\vec{s}}} d_A^{\mathrm{cyc}(C_m\pi)-m} d_B^{\mathrm{cyc}(\pi)-m}. \tag{79}$$

Next, we use the fact (proved in [26]) that for any $\pi \in \mathcal{S}_m$, $\mathrm{cyc}(C_m\pi) + \mathrm{cyc}(\pi) \leq m + 1$ to further bound

$$E_{d_A[d_B]}[\vec{s}] \leq \sum_{\pi\in\mathcal{S}_{\vec{s}}} d_A^{\mathrm{cyc}(C_m\pi)-m} d_B^{1-\mathrm{cyc}(C_m\pi)} = d_B^{1-m} \sum_{\pi\in\mathcal{S}_{\vec{s}}} \left(\frac{d_A}{d_B}\right)^{\mathrm{cyc}(C_m\pi)-m}. \tag{80}$$

On the other hand, if $\mu_1,\ldots,\mu_p$ are the letter frequencies of $\vec{s}$ then (62) and (23) yield

$$E_d[\vec{s}] = \frac{\sum_{\pi\in\mathcal{S}_{\vec{s}}} d^{\mathrm{cyc}(C_m\pi)}}{\prod_{s=1}^p d(d+1)\cdots(d+\mu_s-1)} \geq e^{-\frac{m(m+1)}{2d}} \sum_{\pi\in\mathcal{S}_{\vec{s}}} d^{\mathrm{cyc}(C_m\pi)-m}. \tag{81}$$

Setting $d = d_A/d_B$ and combining (80) and (81) yields the inequality

$$E_{d_A[d_B]}[\vec{s}] \leq E_{d_A/d_B}[\vec{s}]e^{\frac{m(m+1)d_B}{2d_A}}.$$

We then raise both sides to the kth power and sum over $\vec{s}$ to establish the Lemma. $\square$

To avoid lengthy digressions, we avoid presenting any lower bounds for $E_{p,d_A[d_B],k}^m$.

Next, we also consider a model in which some of the random vectors are repeated, which was again first proposed in [23]. Assume that $p^{1/k}$ is an integer. For $s = 1,\ldots,p$ and $j = 1,\ldots,k$, define

$$s^{(j)} := \left\lceil \frac{s}{p^{1-\frac{j}{k}}} \right\rceil.$$

Note that as s ranges from $1,\ldots,p$, $s^{(j)}$ ranges from $1,\ldots,p^{j/k}$. Define $\tilde{M}_{p,d,k} = \sum_{s=1}^p |\tilde{\varphi}_s\rangle\langle\tilde{\varphi}_s|$, where $|\tilde{\varphi}_s\rangle = |\varphi_{s^{(1)}}^1\rangle \otimes\cdots\otimes |\varphi_{s^{(k)}}^k\rangle$. In [23], large-deviation arguments were used to show that for $x = o(1)$, $\|\tilde{M}_{p,d,k}\| = 1 + o(1)$ with high probability. Here we show that this can yield an alternate proof of our main result on the behavior of $\|M_{p,d,k}\|$, at least for small values of x. In particular, we prove

Corollary 12. *For all m, p, d, k,*

$$\tilde{E}^m_{p,d,k} \leq E^m_{p,d,k}.$$

This implies that if $\tilde{\lambda}$ is a randomly drawn eigenvalue of $\tilde{M}_{p,d,k}$, λ is a randomly drawn eigenvalue of $M_{p,d,k}$ and γ is a real number, then $\Pr[\lambda \geq \gamma] \leq \Pr[\tilde{\lambda} \geq \gamma]$. In particular

$$\Pr[\|M_{p,d,k}\| \geq \gamma] \leq d^k \Pr[\|\tilde{M}_{p,d,k}\| \geq \gamma].$$

The proof of Corollary 12 is a direct consequence of the following Lemma, which may be of independent interest.

Lemma 13. *If $s'_i = s'_j$ whenever $s_i = s_j$ for some strings $\vec{s}, \vec{s}' \in [p]^m$ then $E_d[\vec{s}] \leq E_d[\vec{s}']$.*

Proof. The hypothesis of the Lemma can be restated with no loss of generality as saying that $\vec{s}'$ is obtained from $\vec{s}$ by a series of merges, each of which replaces all instances of letters a, b with the letter a. We will prove the inequality for a single such merge. Next, we rearrange $\vec{s}$ so that the a's and b's are at the start of the string. This rearrangement corresponds to a permutation π_0, so we have $E_d[\vec{s}] = \operatorname{tr} \pi_0^\dagger C_m \pi_0 \mathbb{E}[\varphi_a^{\otimes \mu_a} \otimes \varphi_b^{\otimes \mu_b} \otimes \omega]$ and $E_d[\vec{s}'] = \operatorname{tr} \pi_0^\dagger C_m \pi_0 \mathbb{E}[\varphi_a^{\otimes \mu_a + \mu_b} \otimes \omega]$, where ω is a tensor product of various φ_s, with $s \notin \{a, b\}$. Taking the expectation over ω yields a positive linear combination of various permutations, which we absorb into the $\pi_0^\dagger C_m \pi_0$ term by using the cyclic property of the trace. Thus we find

$$E_d[\vec{s}] = \sum_{\pi \in S_m} c_\pi \operatorname{tr} \pi \mathbb{E}[\varphi_a^{\otimes \mu_a} \otimes \varphi_b^{\otimes \mu_b} \otimes I^{m - \mu_a - \mu_b}] \tag{82}$$

$$E_d[\vec{s}'] = \sum_{\pi \in S_m} c_\pi \operatorname{tr} \pi \mathbb{E}[\varphi_a^{\otimes \mu_a + \mu_b} \otimes I^{m - \mu_a - \mu_b}] \tag{83}$$

for some $c_\pi \geq 0$. A single term in the $E_d[\vec{s}]$ sum has the form $c_\pi \mathbb{E}[|\langle \varphi_a | \varphi_b \rangle|^{2f(\pi)}]$ for some $f(\pi) \geq 0$, while for $E_d[\vec{s}']$, the corresponding term is simply c_π. Since $\mathbb{E}[|\langle \varphi_a | \varphi_b \rangle|^{2f(\pi)}] \leq 1$, this establishes the desired inequality. $\square$

IV. APPROACH 3: SCHWINGER-DYSON EQUATIONS

A. Overview

The final method we present uses the Schwinger-Dyson equations[15] to evaluate traces of products of random pure states. First, we show how the expectation of a product of traces may be expressed as an expectation of a similar product involving fewer traces. This will allow us to simplify $E_d[\vec{s}]^k$, and thus to obtain a recurrence relation for $e^m_{p,d,k}$.

B. Expressions involving traces of random matrices

1. Eliminating one φ: Haar random case

We start by considering the case when $k = 1$ (i.e. $|\varphi_i\rangle$ are just Haar-random, without a tensor product structure). Let φ be a density matrix of a Haar-random state over $\mathbb{C}^d$.

Let $A_1, \ldots, A_j$ be matrix-valued random variables that are independent of φ (but there may be dependencies between A_i). We would like to express

$$\mathbb{E}[\operatorname{tr}(\varphi A_1 \varphi A_2 \ldots \varphi A_j)],$$

by an expression that depends only on $A_1, \ldots, A_j$. First, if $\varphi = |\varphi\rangle\langle\varphi|$, then

$$\operatorname{tr}(\varphi A_1 \varphi \ldots \varphi A_i) \operatorname{tr}(\varphi A_{i+1} \varphi \ldots \varphi A_j) = \langle\varphi|A_1 \varphi \ldots \varphi A_i|\varphi\rangle\langle\varphi|A_{i+1}\varphi \ldots \varphi A_j|\varphi\rangle$$
$$= \langle\varphi|A_1 \varphi \ldots A_i \varphi A_{i+1} \ldots \varphi A_j|\varphi\rangle = \operatorname{tr}(\varphi A_1 \ldots \varphi A_j). \tag{84}$$

This allows to merge expressions that involve the same matrix φ.

Second, observe that $\varphi = U|0\rangle\langle 0|U^\dagger$, where U is a random unitary and $|0\rangle$ is a fixed state. By applying eq. (19) from [15], we get

$$\mathbb{E}[\operatorname{tr}(\varphi A_1 \varphi A_2 \ldots \varphi A_j)] = -\frac{1}{d}\sum_{i=1}^{j-1}\mathbb{E}[\operatorname{tr}(\varphi A_1 \ldots \varphi A_i)\operatorname{tr}(\varphi A_{i+1} \ldots \varphi A_j)]$$

$$+\frac{1}{d}\sum_{i=1}^{j}\mathbb{E}[\operatorname{tr}(\varphi A_1 \ldots A_{i-1}\varphi)\operatorname{tr}(A_i \varphi A_{i+1} \ldots \varphi A_j)].$$

Because of (84), we can replace each term in the first sum by $\mathbb{E}[\operatorname{tr}(\varphi A_1 \ldots \varphi A_j)]$. Moving those terms to the left hand side and multiplying everything by $\frac{d}{d+j-1}$ gives

$$\mathbb{E}[\operatorname{tr}(\varphi A_1 \varphi A_2 \ldots \varphi A_j)] = \frac{1}{d+j-1}\sum_{i=1}^{j}\mathbb{E}[\operatorname{tr}(\varphi A_1 \ldots A_{i-1}\varphi)\operatorname{tr}(A_i \varphi A_{i+1} \ldots \varphi A_j)]. \tag{85}$$

For $i = j$, we have

$$\operatorname{tr}(\varphi A_1 \ldots A_{j-1}\varphi)\operatorname{tr}(A_j) = \operatorname{tr}(\varphi A_1 \ldots A_{j-1})\operatorname{tr}(A_j). \tag{86}$$

Here, we have applied $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ and $\varphi^2 = \varphi$. For $i < j$, we can rewrite

$$\operatorname{tr}(\varphi A_1 \ldots A_{i-1}\varphi)\operatorname{tr}(A_i \varphi A_{i+1} \ldots \varphi A_j) = \operatorname{tr}(\varphi A_1 \ldots A_{i-1})\operatorname{tr}(\varphi A_{i+1} \ldots \varphi A_j A_i)$$
$$= \operatorname{tr}(\varphi A_1 \ldots A_{i-1}\varphi A_{i+1} \ldots \varphi A_j A_i). \tag{87}$$

By combining (85), (86) and (87), we have

$$\mathbb{E}[\operatorname{tr}(\varphi A_1 \varphi A_2 \ldots \varphi A_j)] = \frac{1}{d+j-1}\left(\mathbb{E}[\operatorname{tr}(\varphi A_1 \ldots \varphi A_{j-1})\operatorname{tr}(A_j)] + \sum_{i=1}^{j-1}\mathbb{E}[\operatorname{tr}(\varphi A_1 \ldots A_{i-1}\varphi A_{i+1} \ldots \varphi A_j A_i)]\right) \tag{88}$$

$$\leq \frac{1}{d}\left(\mathbb{E}[\operatorname{tr}(\varphi A_1 \ldots \varphi A_{j-1})\operatorname{tr}(A_j)] + \sum_{i=1}^{j-1}\mathbb{E}[\operatorname{tr}(\varphi A_1 \ldots A_{i-1}\varphi A_{i+1} \ldots \varphi A_j A_i)]\right). \tag{89}$$

2. Consequences

Consider $\mathbb{E}[\operatorname{tr}(\varphi_1 \ldots \varphi_m)]$ with φ_i as described in section IV A. Let $Y_1, \ldots, Y_l$ be the different matrix valued random variables that occur among $\varphi_1, \ldots, \varphi_m$. We can use the procedure described above to eliminate all occurrences of Y_1. Then, we can apply it again to eliminate all occurrences of Y_2, …, Y_{l-1}, obtaining an expression that depends only on $\operatorname{tr}(Y_l)$. Since $\operatorname{tr}(Y_l) = 1$, we can then evaluate the expression.

Each application of (88) generates a sum of trace expressions with positive real coefficients. Therefore, the final expression in $\operatorname{tr}(Y_l)$ is also a sum of terms that involve $\operatorname{tr}(Y_l)$ with positive real coefficients. This means that $\mathbb{E}[\operatorname{tr}(\varphi_1 \ldots \varphi_m)]$ is always a positive real.

3. Eliminating one φ: the tensor product case

We claim

Lemma 14. *Let φ be a tensor product of k Haar-random states in d dimensions and $A_1, \ldots, A_j$ be matrix-valued random variables which are independent from φ and whose values are tensor products of matrices in d dimensions. Then,*

$$\mathbb{E}[\operatorname{tr}(\varphi A_1 \varphi A_2 \ldots \varphi A_j)] \leq \frac{1 + j^k d^{-1/k}}{d}\mathbb{E}[\operatorname{tr}(\varphi A_1 \ldots \varphi A_{j-1})\operatorname{tr}(A_j)] + \frac{j^k}{d^{1/k}}\sum_{i=1}^{j-1}\mathbb{E}[\operatorname{tr}(\varphi A_1 \ldots A_{i-1}\varphi A_{i+1} \ldots \varphi A_j A_i)].$$

Proof. Because of the tensor product structure, we can express

$$\varphi = \varphi^1 \otimes \varphi^2 \otimes \ldots \varphi^k,$$

$$A_i = A_i^1 \otimes A_i^2 \otimes \ldots A_i^k.$$

We have

$$\mathbb{E}[\operatorname{tr}(\varphi A_1 \varphi A_2 \ldots \varphi A_j)] = \prod_{l=1}^{k} \mathbb{E}[\operatorname{tr}(\varphi^l A_1^l \varphi^l \ldots \varphi^l A_j^l)].$$

We expand each of terms in the product according to (89). Let $C_0 = \mathbb{E}[\operatorname{tr}(\varphi^l A_1^l \ldots \varphi^l A_{j-1}^l) \operatorname{tr}(A_j^l)]$ and

$$C_i = \mathbb{E}[\operatorname{tr}(\varphi^l A_1^l \ldots A_{i-1}^l \varphi^l A_{i+1}^l \ldots \varphi^l A_j^l A_i^l)]$$

for $i \in \{1, 2, \ldots, j-1\}$. (Since each of k subsystems has equal dimension d and are identically distributed, the expectations $C_0, \ldots, C_{j-1}$ are independent of l.) Then, from (89), we get

$$\mathbb{E}[\operatorname{tr}(\varphi A_1 \varphi A_2 \ldots \varphi A_j)] \le \frac{1}{d^k} \prod_{l=1}^{k} (C_0 + C_1 + \ldots + C_{j-1}) = \frac{1}{d^k} \sum_{i_1=0}^{j-1} \ldots \sum_{i_k=0}^{j-1} C_{i_1} \cdots C_{i_k}.$$

Consider one term in this sum. Let r be the number of l for which $i_l = 0$. We apply the arithmetic-geometric mean inequality

$$\frac{x_1 + x_2 + \cdots + x_k}{k} \ge \sqrt[k]{x_1 x_2 \cdots x_k}$$

to

$$x_l = \begin{cases} d^{-\frac{1}{k}} (C_{i_l})^k & \text{if } i_l = 0 \\ d^{\frac{r}{(k-r)k}} (C_{i_l})^k & \text{if } i_l \ne 0 \end{cases}.$$

(In cases if $r = 0$ or $r = k$, we just define $x_l = C_{i_l}$ for all $l \in \{1, 2, \ldots, k\}$.) We now upper-bound the coefficients of $(C_0)^k$ in the resulting sum. For $(C_0)^k$, we have a contribution of 1 from the term which has $i_1 = \ldots = i_k = 0$ and a contribution of at most $d^{-1/k}$ from every other term. Since there are at most j^k terms, the coefficient of $(C_0)^k$ is at most

$$1 + j^k d^{-1/k}.$$

The coefficient of $(C_j)^k$ in each term is at most $d^{\frac{r}{(k-r)k}}$. Since $r \le k-1$ (because the $r = k$ terms only contain C_0's), we have $d^{\frac{r}{(k-r)k}} \le d^{\frac{k-1}{k}}$. The Lemma now follows from there being at most j^k terms. $\square$

C. Main results

1. Haar random case

We would like to upper-bound

$$e_{p,d,1}^m = \frac{1}{d} \sum_{s_1=1}^{p} \ldots \sum_{s_m=1}^{p} \mathbb{E}[\operatorname{tr}(\varphi_{s_1} \ldots \varphi_{s_m})].$$

Lemma 15.

$$e_{p,d,1}^m \le \sum_{l=0}^{m-2} e_{p,d,1}^l e_{p,d,1}^{m-l-1} + \frac{p+m^3}{d} e_{p,d,1}^{m-1}. \tag{90}$$

Proof. In section IV D 1. □

Using $e^0_{p,d,1} = \mathrm{tr}(I)/d = 1$, we can state Lemma 15 equivalently as

$$e^m_{p,d,1} \le \sum_{l=0}^{m-1} e^l_{p,d,1} e^{m-l-1}_{p,d,1} + \left(\frac{p+m^3}{d} - 1\right) e^{m-1}_{p,d,1}. \tag{91}$$

Define $\tilde{x} = (p+m^3)/d$ (and note that it is not exactly the same as the variable of the same name in Section II). Then (91) matches the recurrence for the Narayana coefficients in (54). Thus we have

Corollary 16.

$$e^m_{p,d,1} \le \sum_{\ell=1}^{m} N(m,\ell)\tilde{x}^\ell = \beta_m(\tilde{x}) \le (1+\sqrt{\tilde{x}})^{2m} \tag{92}$$

Similar arguments (which we omit) establish the lower bound $e^m_{p,d,1} \ge \sum_\ell N(m,\ell)(p)_\ell/(d+m)^\ell$, which is only slightly weaker than the bound stated in Theorem 1 and proved in Lemma 8.

2. Tensor product case

The counterpart of Lemma 15 is

Lemma 17.

$$e^m_{p,d,k} \le \left(1+\frac{m^k}{d^{1/k}}\right)\sum_{l=0}^{m-2} e^l_{p,d,k} e^{m-l-1}_{p,d,k} + \left(\frac{p}{d^k} + \frac{3m^{k+3}}{d^{1/k}}\right) e^{m-1}_{p,d,k} \tag{93}$$

$$= \left(1+\frac{m^k}{d^{1/k}}\right)\sum_{l=0}^{m-1} e^l_{p,d,k} e^{m-l-1}_{p,d,k} + \left(\frac{p}{d^k} + 3\frac{m^{k+3}}{d^{1/k}} - \left(1+\frac{m^k}{d^{1/k}}\right)\right) e^{m-1}_{p,d,k} \tag{94}$$

$$\tag{95}$$

This time we set $\tilde{x}_k = \frac{p}{d^k} + 3\frac{m^{k+3}}{d^{1/k}}$. Also define $\gamma = m^k/d^{1/k}$. Then Lemma 17 implies that $e^m_{p,d,k} \le (1+\gamma)^m[y^m]\tilde{F}(\tilde{x}_k, y)$, where $\tilde{F}$ satisfies the recurrence

$$\tilde{F} = 1 + y\tilde{F}^2 + y\left(\frac{\tilde{x}_k}{1+\gamma} - 1\right)\tilde{F}. \tag{96}$$

Thus we obtain

Corollary 18.

$$e^m_{p,d,k} \le (1+\gamma)^m \beta_m\left(\frac{\tilde{x}_k}{1+\gamma}\right) \tag{97}$$

$$\le \left(\frac{\tilde{x}_k}{x}\right)^m \beta_m(x) \le \exp\left(\frac{3m^{k+4}}{xd^{1/k}}\right)\beta_m(x) \tag{98}$$

Proof. (97) follows from the preceding discussion as well as the relation between β_m and the recurrence (96), which was discussed in Section III A and in [30, 31]. The first inequality in (98) is because $\beta_m(x(1+\epsilon)) \le (1+\epsilon)^m\beta_m(x)$ for any $\epsilon \ge 0$, which in turn follows from the fact that $\beta_m(x)$ is a degree-m polynomial in x with nonnegative coefficients. The second inequality follows from the inequality $1+\epsilon \le e^\epsilon$. □

D. Proofs

1. Proof of Lemma 15

We divide the terms $\mathbb{E}[\mathrm{tr}(\varphi_{s_1} \ldots \varphi_{s_m})]$ into several types.

First, we consider terms for which $s_1 \notin \{s_2, \ldots, s_m\}$. Then, φ_{s_1} is independent from $\varphi_{s_2} \ldots \varphi_{s_m}$. Because of linearity of expectation , we have

$$\mathbb{E}[\mathrm{tr}(\varphi_{s_1} \ldots \varphi_{s_m})] = \mathrm{tr}(\mathbb{E}[\varphi_{s_1}]\mathbb{E}[\varphi_{s_2} \ldots \varphi_{s_m}]) = \mathrm{tr}\left(\frac{I}{d}\mathbb{E}[\varphi_{s_2} \ldots \varphi_{s_m}]\right) = \frac{1}{d}\mathbb{E}[\mathrm{tr}(\varphi_{s_2} \ldots \varphi_{s_m})].$$

By summing over all possible $s_1 \in [p]$, the sum of all terms of this type is $\frac{p}{d}$ times the sum of all possible $\mathbb{E}[\mathrm{tr}(\varphi_{s_2} \ldots \varphi_{s_m})]$ with $s_1 \notin \{s_2, \ldots, s_m\}$, i.e., $\frac{p}{d}$ times $E_{p-1,d,1}^{m-1}$.

For the other terms, we can express them as

$$\mathbb{E}[\mathrm{tr}(\varphi_{s_1} Y_1 \varphi_{s_1} Y_2 \ldots \varphi_{s_1} Y_j)] \tag{99}$$

with $Y_1, \ldots, Y_j$ being products of φ_i for $i \neq s_1$. (Some of those products may be empty, i.e. equal to I.)

To simplify the notation, we denote $\varphi = \varphi_{s_1}$. Because of (89), (99) is less than or equal to

$$\frac{1}{d}\left(\sum_{i=1}^{j-1}\mathbb{E}[\mathrm{tr}(\varphi Y_1 \varphi \ldots Y_{i-1} Y_i \varphi \ldots \varphi Y_j] + \mathbb{E}[\mathrm{tr}(\varphi Y_1 \varphi Y_2 \ldots \varphi Y_{j-1})\,\mathrm{tr}(Y_j)]\right) \tag{100}$$

We handle each of the two terms in (100) separately. For each the term in the sum, we will upper-bound the sum of them all (over all $\mathbb{E}[\mathrm{tr}(\varphi Y_1 \varphi Y_2 \ldots \varphi Y_j)]$) by $\frac{1}{d}E_{p,d,1}^{m-1}$ times the maximum number of times the same term can appear in the sum.

Therefore, we have to answer the question: given a term $\mathbb{E}[\mathrm{tr}(Z_1 \ldots Z_{m-1})]$, what is the maximum number of ways how this term can be generated as $\mathbb{E}[\mathrm{tr}(\varphi Y_1 \varphi \ldots Y_{i-1} Y_i \varphi \ldots \varphi Y_j]$?

Observe that $\varphi = Z_1$. Thus, given $Z_1 \ldots Z_{m-1}$, φ is uniquely determined. Furthermore, there are at most m locations in $Z_1 \ldots Z_{m-1}$ which could be the boundary between Y_{i-1} and Y_i. The original term $\mathbb{E}[\mathrm{tr}(\varphi Y_1 \ldots \varphi Y_i)]$ can then be recovered by adding φ in that location. Thus, each term can be generated in at most m ways and the sum of them all is at most $\frac{m}{d}E_{p,d,1}^{m-1}$.

It remains to handle the terms of the form

$$\mathbb{E}[\mathrm{tr}(\varphi Y_1 \varphi Y_2 \ldots \varphi Y_{j-1})\,\mathrm{tr}(Y_j)]. \tag{101}$$

We consider two cases:

Case 1: There is no φ_i which occurs both in Y_j and in at least one of $Y_1, \ldots, Y_{j-1}$. Then, the matrix valued random variables $\varphi Y_1 \varphi Y_2 \ldots \varphi Y_{j-1}$ and Y_j are independent. Therefore, we can rewrite (101) as

$$\mathbb{E}[\mathrm{tr}(\varphi Y_1 \varphi Y_2 \ldots \varphi Y_{j-1})]\mathbb{E}[\mathrm{tr}(Y_j)]. \tag{102}$$

Fix $Y_1, \ldots, Y_{j-1}$. Let l be the length of Y_j and let o be the number of different φ_i that occur in $Y_1 \ldots Y_{j-1}$. Then, there are $p - o - 1$ different φ_is which can occur in Y_j (i.e., all p possible φ_is, except for φ_{s_1} and those o which occur in $Y_1 \ldots Y_{j-1}$).

Therefore, the sum of $\mathbb{E}[\mathrm{tr}(Y_j)]$ over all possible Y_j is exactly $E_{p-o-1,d,1}^l$. We have $E_{p-t,d,1}^l \leq E_{p-o-1,d,1}^l \leq E_{p,d,1}^l$. Therefore, the sum of all terms (102) in which Y_j is of length l is lower-bounded by the sum of all

$$E_{p-t,d,1}^l\mathbb{E}[\mathrm{tr}(\varphi Y_1 \varphi Y_2 \ldots \varphi Y_{j-1})]$$

which is equal to $E_{p-t,d,1}^l E_{p,d,1}^{m-l-1}$. Similarly, it is upper-bounded by $E_{p,d,1}^l E_{p,d,1}^{m-l-1}$.

Case 2: There exists φ_i which occurs both in Y_j and in some Y_l, $l \in \{1, \ldots, j-1\}$.

We express $Y_j = Z\varphi_i W$ and $Y_l = Z'\varphi_i W'$. Then, (101) is equal to

$$\mathbb{E}[\mathrm{tr}(\varphi Y_1 \ldots Y_{l-1}\varphi Z\varphi_i W\varphi Y_{l+1} \ldots \varphi Y_{j-1})\,\mathrm{tr}(Z'\varphi_i W')] =$$

$$\mathbb{E}[\mathrm{tr}(\varphi Y_1 \ldots Y_{l-1}\varphi Z\varphi_i W'Z'\varphi_i W\varphi Y_{l+1} \ldots \varphi Y_{j-1})].$$

In how many different ways could this give us the same term $\mathbb{E}[\mathrm{tr}(Z_1 \ldots Z_{m-1})]$?

Given $Z_1, \ldots, Z_{m-1}$, we know $\varphi = Z_1$. Furthermore, we can recover Y_j by specifying the location of the first φ_i, the second φ_i and the location where W' ends and Z' begins. There are at most $m-1$ choices for each of those three parameters. Once we specify them all, we can recover the original term (101). Therefore, the sum of all terms (101) in this case is at most $(m-1)^3$ times the sum of all $\mathbb{E}[\mathrm{tr}(Z_1 \ldots Z_{m-1})]$, which is equal to $E_{p,d,1}^{m-1}$.

Overall, we get

$$E_{p,d,1}^m \leq \frac{p}{d} E_{p-1,d,1}^{m-1} + \frac{m}{d} E_{p,d,1}^{m-1} + \sum_{l=0}^{m-2} E_{p,d,1}^l E_{p,d,1}^{m-l-1} + \frac{(m-1)^3}{d} E_{p,d,1}^{m-1}, \tag{103}$$

with the first term coming from the terms where $s_1 \notin \{s_2, \ldots, s_k\}$, the second term coming from the bound on the sum in (100) and the third and the fourth terms coming from Cases 1 and 2. By combining the terms, we can rewrite (103) as

$$E_{p,d,1}^m \leq \frac{p+m^3}{d} E_{p,d,1}^{m-1} + \frac{1}{d} \sum_{l=0}^{m-2} E_{p,d,1}^l E_{p,d,1}^{m-l-1}. \tag{104}$$

Dividing (104) by d completes the proof.

We remark as well that these techniques can yield a lower bound for $E_{p,d,1}$. To do so, we apply the inequality $1/(d+j-1) \geq 1/(d+m)$ to (88), and then combine the lower bounds from the $s_1 \notin \{s_2, \ldots, s_k\}$ case and Case 1. (For the other cases, we can use 0 as the lower bound, since we know that the expectation of any product of traces is positive.) This yields

$$\frac{d}{d+m} \left(\sum_{l=0}^{m-2} e_{p,d,1}^l e_{p-t,d,1}^{m-l-1} + \frac{p}{d} e_{p-1,d,1}^{m-1} \right) \leq e_{p,d,1}^m. \tag{105}$$

2. Proof of Lemma 17

The proof is the same as for Lemma 17, except that, instead of (89) we use Lemma 14.

The first term in (103), $\frac{p}{d^k} E_{p-1,d,k}^{m-1}$, remains unchanged. The terms $\mathbb{E}[\mathrm{tr}(\varphi Y_1 \ldots Y_{i-1} Y_i \varphi \ldots \varphi Y_j)]$ in (100) are now multiplied by $\frac{j^k}{d^{1/k}}$ instead of $\frac{1}{d}$. We have $\frac{j^k}{d^{1/k}} \leq \frac{m^k}{d^{1/k}}$. Therefore, the second term in (103) changes from $\frac{m}{d} E_{p,d,1}^{m-1}$ to $\frac{m^{k+1}}{d^{1/k}} E_{p,d,k}^{m-1}$.

The terms $\mathbb{E}[\mathrm{tr}(\varphi Y_1 \ldots \varphi Y_{j-1}) \mathrm{tr}(Y_j)]$ in (100) acquire an additional factor of $1 + \frac{j^k}{d^{1/k}} \leq 1 + \frac{m^k}{d^{1/k}}$. This factor is then acquired by the third and the fourth terms in (103). Thus, we get

$$E_{p,d,k}^m \leq \frac{p}{d^k} E_{p-1,d,k}^{m-1} + \frac{m^{k+1}}{d^{1/k}} E_{p,d,k}^{m-1} + \left(1 + \frac{m^k}{d^{1/k}}\right) \sum_{l=0}^{m-2} E_{p,d,k}^l E_{p,d,1}^{m-l-1} + \left(1 + \frac{m^k}{d^{1/k}}\right) \frac{(m-1)^3}{d} E_{p,d,k}^{m-1}.$$

The lemma now follows from merging the second term with the fourth term.

E. Relation to combinatorial approach

The recursive approach of this section appears on its face to be quite different from the diagrammatic and combinatorial methods discussed earlier. However, the key recursive step in (85) (or equivalently (88)) can be interpreted in terms of the sorts of sums over permutations seen in Section III.

Consider an expression of the form $X = \mathbb{E}[\mathrm{tr}(\varphi A_1 \varphi A_2 \ldots \varphi A_j)]$. For the purposes of this argument, we will ignore the fact that $A_1, \ldots, A_j$ are random variables. Letting C_j denote the j-cycle, we can rewrite X as

$$X = \mathrm{tr}(C_j \mathbb{E}[\varphi^{\otimes j}](A_1 \otimes A_2 \otimes \cdots \otimes A_j)) = \mathrm{tr}(\mathbb{E}[\varphi^{\otimes j}](A_1 \otimes A_2 \otimes \cdots \otimes A_j)),$$

since $C_j |\varphi\rangle^{\otimes j} = |\varphi\rangle^{\otimes j}$. Next we apply (61) and obtain

$$X = \frac{\sum_{\pi \in \mathcal{S}_j} \mathrm{tr}(\pi(A_1 \otimes A_2 \otimes \cdots \otimes A_j))}{d(d+1) \cdots (d+j-1)}.$$

We will depart here from the approach in Section III by rewriting the sum over $\mathcal{S}_j$. For $1 \leq i \leq j$, let (i, j) denote the permutation that exchanges positions i and j, with $(j, j) = e$ standing for the identity permutation. We also define $\mathcal{S}_{j-1} \subset \mathcal{S}_j$ to be the subgroup of permutations of the first $j-1$ positions. Since $(1, j), \ldots, (j-1, j), (j, j)$ are

a complete set of coset representatives for $\mathcal{S}_{j-1}$, it follows that any $\pi \in \mathcal{S}_j$ can be uniquely expressed in the form $(i,j)\pi'$ with $1 \leq j$ and $\pi' \in \mathcal{S}_{j-1}$. Our expression for X then becomes

$$
\begin{aligned}
X &= \frac{1}{d+j-1} \operatorname{tr}\left(\sum_{i=1}^{j}(i,j)\frac{\sum_{\pi'\in\mathcal{S}_{j-1}}\pi'(A_1 \otimes A_2 \otimes \cdots \otimes A_j)}{d(d+1)\cdots(d+j-2)}\right).\\
&= \frac{1}{d+j-1}\mathbb{E}\left[\operatorname{tr}\left(\sum_{i=1}^{j}(i,j)(\varphi^{\otimes j-1} \otimes I)(A_1 \otimes A_2 \otimes \cdots \otimes A_j)\right)\right]\\
&= \frac{1}{d+j-1}\mathbb{E}\left[\operatorname{tr}\left(\varphi A_1 \varphi A_2 \cdots \varphi A_{j-1}\right)\operatorname{tr}(A_j) + \sum_{i=1}^{j-1}\operatorname{tr}(\varphi A_1)\cdots\operatorname{tr}(\varphi A_{i-1})\operatorname{tr}(A_j \varphi A_i)\operatorname{tr}(\varphi A_{i+1})\cdots\operatorname{tr}(\varphi A_{j-1})\right],
\end{aligned}
$$

which matches the expression in (85), or equivalently, (88).

The difference in approaches can then be seen as stemming from the different ways of summing over $\pi \in \mathcal{S}_j$. In Section III (and to some extent, Section II), we analyzed the entire sum by identifying leading-order terms and deriving a perturbative expansion that accounted for all the other terms. By contrast, the approach of this section is based on reducing the sum over $\mathcal{S}_j$ to a similar sum over $\mathcal{S}_{j-1}$.

V. LOWER BOUNDS ON THE SPECTRUM

The bulk of our paper has been concerned with showing that $\|M\|$ is unlikely to be too large (Corollary 5). Since we give asymptotically sharp bounds on $d^{-k}\mathbb{E}[\operatorname{tr} M^m]$, we in fact obtain asymptotically convergent estimates of the eigenvalue density of M (Corollary 6). However, this does not rule out the possibility that a single eigenvalue of M might be smaller than $(1-\sqrt{x})^2$; rather, it states that the expected number of such eigenvalues is $o(d^k)$.

In fact, our method was successful in proving asymptotically sharp estimates on the largest eigenvalue of M. We now turn to proving bounds on the smallest eigenvalue of M. To use the trace method to show that w.h.p. there are *no* small eigenvalues, one would like to upper bound expressions such as $\mathbb{E}[\operatorname{tr}(M-\lambda I)^{2m}]$, for an appropriate choice of λ. If we succeed in bounding such an expression then the $\lambda_{\min}$ (the smallest eigenvalue of M) is lower bounded by

$$
\mathbb{E}[(\lambda - \lambda_{\min})^2] \leq \left(\mathbb{E}[\operatorname{tr}(M-\lambda I)^{2m}]\right)^{1/m}, \tag{106}
$$

and hence

$$
\mathbb{E}[\lambda_{\min}] \geq \lambda - \left(\mathbb{E}[\operatorname{tr}(M-\lambda I)^{2m}]\right)^{1/2m}. \tag{107}
$$

Let us first describe a failed attempt to bound this result, before giving the correct approach. To bound $\mathbb{E}[\operatorname{tr}(M-\lambda I)^{2m}]$, the natural first attempt is to use the expansion

$$
\mathbb{E}[\operatorname{tr}(M-\lambda I)^{2m}] = \sum_{n=0}^{2m}\binom{2m}{n}\mathbb{E}[\operatorname{tr}(M^n)]\left(-\lambda\right)^{2m-n}. \tag{108}
$$

One might then attempt to estimate each term in the above expansion in turn. Unfortunately, what happens is the following: the leading order (rainbow) terms for $\mathbb{E}[\operatorname{tr}(M^n)]$ can be summed directly over n. One may show that this sum contributes a result to $\mathbb{E}[\operatorname{tr}(M-\lambda I)^{2m}]$ which grows roughly as $\max\{((\sqrt{x}-1)^2-\lambda)^{2m},((\sqrt{x}+1)^2-\lambda)^{2m}\}$. That is, it is dominated by either the largest or smallest eigenvalue of the limiting distribution, depending on the value of λ. However, we are unable to control the corrections to this result. While they are suppressed in powers of $1/d$, they grow rapidly with m due to the binomial factor, causing this attempt to fail.

We now describe a simple alternate approach. Let us work within the Feynman diagram framework. By (15), the spectrum of $M_{p,d,k}$ is close to that of $\hat{M}_{p,d,k}$ with high probability, so we can translate bounds on $\lambda_{\min}$ in the Gaussian ensemble to bounds on the smallest eigenvalue in the normalized ensemble.

Having reduced to the Gaussian ensemble, we now construct a diagrammatic series for $\mathbb{E}[\operatorname{tr}(\hat{M}-\lambda I)^{2m}]$. One way to construct such a diagrammatic series is to add in extra diagrams, in which rather than having m pairs of vertices, we instead have n pairs of vertices, interspersed with $m-n$ "identity operators", where nothing happens: the solid lines simply proceed straight through. However, there already is a particular contraction in our existing diagrammatic series in which solid lines proceed straight through. This is a particular contraction of neighboring vertices, in which

a dashed line connects the two vertices and *all* vertical lines leaving the two vertices are connected to each other. So, we can obtain the same result by using our original diagrammatic expansion, but with a change in the rules for weighting diagrams. If a a diagram has a certain number, c, of pairs of neighboring vertices contracted in the given way, then we adjust the weight of the diagram by

$$\Big(\frac{d^{-k}p - \lambda}{d^{-k}p}\Big)^c = \frac{x - \lambda}{x}. \tag{109}$$

If $d^{-k}p - \lambda \geq 0$, then this new series consists only of positive terms and we can use our previous techniques for estimating the series, bounding it by the sum of rainbow diagrams, plus higher order corrections. The sum of rainbow diagrams changes in this approach. One could use a new set of generating functionals to evaluate the new sum of rainbow diagrams, but we can in fact find the result more directly: we can directly use the fact that this sum is bounded by $d^k \max\{((\sqrt{x}-1)^2 - \lambda)^{2m}, ((\sqrt{x}+1)^2 - \lambda)^{2m}\}$. The corrections remain small. Taking the smallest value of λ such that $x - \lambda \geq 0$, we have $\lambda = x$, and so we find that, for $x > 1$, the sum of these diagrams is bounded by $d^k(2\sqrt{x}+1)^{2m}$. This gives us a bound that, for any $\epsilon > 0$, the expectation value for the smallest eigenvalue is asymptotically greater than

$$x - 2\sqrt{x} - 1 - \epsilon, \tag{110}$$

and hence using concentration of measure arguments and the above reduction to the Gaussian ensemble, we can then show that, for any $\epsilon > 0$, with high probability, the smallest eigenvalue of a matrix chosen randomly from the uniform ensemble is greater than or equal to $x - 2\sqrt{x} - 1 - \epsilon$.

On the other hand, if $x < 1$, then we will need to instead consider $\mathbb{E}[\mathrm{tr}(\hat{M}' - \lambda I)^{2m}]$ where $\hat{M}'$ is the Gram matrix of the ensemble. Since $\hat{M}'$ has the same spectrum as $\hat{M}$ but is only $p \times p$, all of the terms in (108) are identical except that $\mathrm{tr}\, I$ equals p instead of d^k. We can use a similar diagrammatic technique to incorporating the identity terms. Now each term of $\hat{M}$ contributes the pair of vertices from Fig. 2(a), but in the opposite order. Along the horizontal, the solid lines are the internal lines and the dashed lines are external. Now the identity diagrams correspond to the case when the *dashed* lines proceed straight through. These components of a diagram initially had a contribution of 1 (with k closed solid loops canceling the natural d^{-k} contribution from each pair of vertices). Thus, adding in the $-\lambda I$ terms results in a multiplicative factor of $(1 - \lambda)$ for each vertex pair with the configuration where the dashed lines go straight through. Now we can choose λ to be as large as 1 and still have each diagram be nonnegative. The resulting bound on $\mathbb{E}[\mathrm{tr}(\hat{M}' - \lambda I)^{2m}]$ is $p \max\{((\sqrt{x}-1)^2 - 1)^{2m}, ((\sqrt{x}+1)^2 - 1)^{2m}\}$ plus small corrections. We find that the smallest eigenvalue is $\geq 1 - 2\sqrt{x} - x - \epsilon$ with high probability.

Combining these bounds, we find that the smallest eigenvalue is asymptotically no lower than $(1 - \sqrt{x})^2 - 2\min(1, x)$. This is within a $1 - o(1)$ factor of the unproven-but-true value of $(1 - \sqrt{x})^2$ in the limits $x \to 0$ and $x \to \infty$.

We believe that it should be possible to improve this result to get an asymptotic lower bound of $(1 - \sqrt{x})^2$, staying within the framework of trace methods, using any of the three techniques we have used. This will require a more careful estimate of the negative terms to show that our methods remain valid. We leave the solution of this problem to future work.

[1] A. Abeyesinghe, I. Devetak, P. Hayden, and A. Winter. The mother of all protocols: Restructuring quantum information's family tree. *Proc. Roc. Soc. A*, 465(2108):2537–2563, 2009. arXiv:quant-ph/0606225.

[2] R. Adamczak, A. Litvak, A. Pajor, and N. Tomczak-Jaegermann. Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles. *J. Amer. Math. Soc.*, Oct 2009. arXiv:0903.2323.

[3] R. Adamczak, A. E. Litvak, A. Pajor, and N. Tomczak-Jaegermann. Restricted isometry property of matrices with independent columns and neighborly polytopes by random sampling, 2009. arXiv:0904.4723.

[4] R. Ahlswede and A. Winter. Strong converse for identification via quantum channels. *IEEE Trans. Inf. Theory*, 48(3):569–579, 2002. arXiv:quant-ph/0012127.

[5] A. Anderson, R. C. Meyrs, and V. Periwal. Complex random surfaces. *Phys. Lett. B*, 254(1-2):89 – 93, 1991.

[6] A. Anderson, R. C. Myers, and V. Periwal. Branched polymers from a double-scaling limit of matrix models. *Nuclear Physics B*, 360(2-3):463 – 479, 1991.

[7] R. Arratia, B. Bollobás, and G. Sorkin. The interlace polynomial of a graph. *J. Comb. Th. B,*, 92(2):199–233,, 2004.

[8] A. Ben-Aroya, O. Schwartz, and A. Ta-Shma. Quantum expanders: motivation and construction. In *CCC*, 2008. arXiv:0709.0911 and arXiv:quant-ph/0702129.

[9] C. H. Bennett, P. Hayden, D. W. Leung, P. W. Shor, and A. J. Winter. Remote preparation of quantum states. *ieeeit*, 51(1):56–74, 2005. quant-ph/0307100.

[10] A. Bose and A. Sen. Another look at the moment method for large dimensional random matrices. *Elec. J. of Prob.*, 13(21):588–628, 2008.

[11] M. Christandl. *The structure of bipartite quantum states: Insights from group theory and cryptography.* PhD thesis, University of Cambridge, 2006. arXiv:quant-ph/0604183.

[12] J. Feinberg and A. Zee. Renormalizing rectangles and other topics in random matrix theory. *J. Stat. Phys.*, 87(3–4):473–504, 1997.

[13] P. Forrester. Log-gases and random matrices. unpublished manuscript. Chapter 2. `http://www.ms.unimelb.edu.au/∼matpjf/matpjf.html`.

[14] M. B. Hastings. Entropy and entanglement in quantum ground states. *Phys. Rev. B*, 76:035114, 2007. arXiv:cond-mat/0701055.

[15] M. B. Hastings. Random unitaries give quantum expanders. *Phys. Rev. A*, 76:032315, 2007. arXiv:0706.0556.

[16] M. B. Hastings. A counterexample to additivity of minimum output entropy. *Nature Physics*, 5, 2009. arXiv:0809.3972.

[17] P. Hayden, P. W. S. D. W. Leung, and A. J. Winter. Randomizing quantum states: Constructions and applications. *Comm. Math. Phys.*, 250:371–391, 2004. arXiv:quant-ph/0307104.

[18] P. Hayden, D. W. Leung, and A. Winter. Aspects of generic entanglement. *Comm. Math. Phys.*, 265:95, 2006. arXiv:quant-ph/0407049.

[19] P. Hayden and A. J. Winter. Counterexamples to the maximal p-norm multiplicativity conjecture for all $p > 1$. *Comm. Math. Phys.*, 284(1):263–280, 2008. arXiv:0807.4753.

[20] R. A. Horn and C. Johnson. *Matrix Analysis.* Cambridge University Press, 1985.

[21] I. M. Johnstone. On the distribution of the largest eigenvalue in principle components analysis. *Annals of Statistics*, 29(2):295–327, 2001.

[22] M. Ledoux. *The Concentration of Measure Phenomenon.* AMS, 2001.

[23] D. W. Leung and A. J. Winter. Locking 2-LOCC distillable common randomness and LOCC-accessible information. in preparation.

[24] A. Montanaro. On the distinguishability of random quantum states. *Comm. Math. Phys.*, 273(3):619–636, 2007. arXiv:quant-ph/0607011v2.

[25] R. C. Myers and V. Periwal. From polymers to quantum gravity: Triple-scaling in rectangular random matrix models. *Nuclear Physics B*, 390(3):716 – 746, 1993.

[26] A. Nica and R. Speicher. *Lectures on the Combinatorics of Free Probability.* Cambridge University Press, 2006.

[27] M. Rudelson. Random vectors in the isotropic position. *J. Func. Anal.*, 164(1):60–72, 1999.

[28] G. Smith and J. Smolin. Extensive nonadditivity of privacy. arXiv:0904.4050, 2009.

[29] R. Speicher. Free probability theory and non-crossing partitions. *Lothar. Comb*, B39c, 1997.

[30] R. P. Stanley. *Enumerative Combinatorics, vol. 2.* Cambridge University Press, 1999. Exercise 6.36 and references therein.

[31] R. A. Sulanke. The Narayana distribution. *J. of Stat. Planning and Inference*, 101(1–2):311–326, 2002.

[32] J. Verbaarschot. Spectrum of the QCD Dirac operator and chiral random matrix theory. *Phys. Rev. Lett.*, 72(16):2531–2533, Apr 1994.

[33] J. Verbaarschot. The spectrum of the Dirac operator near zero virtuality for $N_c = 2$ and chiral random matrix theory. *Nuclear Physics B*, 426(3):559 – 574, 1994.

[34] J. Yard and I. Devetak. Optimal quantum source coding with quantum information at the encoder and decoder, 2007. arXiv:0706.2907.